

FUNDAÇÃO GETULIO VARGAS
ESCOLA BRASILEIRA DE ECONOMIA E FINANÇAS

DIEGO ASENSI TAVARES MARCELO

**PARAMETRIZAÇÃO DE DESPESAS MUNICIPAIS E DETECÇÃO DE
ANOMALIAS**

Rio de Janeiro

2021

DIEGO ASENSI TAVARES MARCELO

**PARAMETRIZAÇÃO DE DESPESAS MUNICIPAIS E DETECÇÃO DE
ANOMALIAS**

Dissertação para obtenção do grau de mestre apresentada à Escola
de Pós-Graduação em Economia da Fundação Getúlio Vargas.

Área de Concentração: Ciência de Dados Aplicada à Economia.

Orientador: Genaro Dueire Lins

Rio de Janeiro

2021

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada pelo Sistema de Bibliotecas/FGV

Marcelo, Diego Asensi Tavares

Parametrização de despesas municipais e detecção de anomalias / Diego Asensi Tavares Marcelo. – 2021.
75 f.

Dissertação (mestrado) - Fundação Getulio Vargas, Escola Brasileira de Economia e Finanças.

Orientador: Genaro Dueire Lins.

Inclui bibliografia.

1. Corrupção administrativa. 2. Auditoria - Processamento de dados. 3. São Paulo (Estado) - Municípios. 4. Administração municipal. I. Lins, Genaro Dueire. II. Fundação Getulio Vargas. Escola Brasileira de Economia e Finanças. III. Título.

CDD – 352.16

DIEGO ASENSI TAVARES MARCELO

"PARAMETRIZAÇÃO DE DESPESAS MUNICIPAIS E DETECÇÃO DE ANOMALIAS".

Dissertação apresentado(a) ao Curso de Mestrado em Economia do(a) EPGE Escola Brasileira de Economia e Finanças - FGV EPGE para obtenção do grau de Mestre(a) em Economia Empresarial e Finanças.

Data da defesa: 24/05/2021

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

Presidente da Comissão Examinadora: Prof^o/a Genaro Dueire Lins


Genaro Dueire Lins
Orientador




Carlos Eugênio Ellery Lustosa da Costa
Membro

por 

Gerson Luis Romantini
Membro

por 

Carlos Emmanuel Joppert Ragazzo
Membro

por 

Nos termos da Lei nº 13.979 de 06/02/20 - DOU nº 27 de 07/02/20 e Portaria MEC nº 544 de 16/06/20 - DOU nº 114 de 17/06/20 que dispõem sobre a suspensão temporária das atividades acadêmicas presenciais e a utilização de recursos tecnológicos face ao COVID-19, as apresentações das defesas de Tese e Dissertação, de forma excepcional, serão realizadas de forma remota e síncrona, incluindo-se nessa modalidade membros da banca e discente.



Ricardo de Oliveira Cavalcanti
Coordenador



Antonio de Araujo Freitas Junior
Pró-Reitor de Ensino, Pesquisa e Pós-Graduação FGV

Antonio Freitas, PhD
Pró-Reitor de Ensino, Pesquisa e Pós-Graduação
Fundação Getúlio Vargas

Instrução Normativa nº 01/19, de 09/07/19 - Pró-Reitoria FGV

Em caso de participação de Membro(s) da Banca Examinadora de forma não-presencial*, o Presidente da Comissão Examinadora assinará o documento como representante legal, delegado por esta I.N.

*Skype, Videoconferência, Apps de vídeo etc

RESUMO

O presente estudo, após realizar uma revisão da literatura de métodos de correção de corrupção, e expor os modelos microeconômicos clássicos para sua análise, propõe a utilização de métodos de aprendizado de máquina para a parametrização de despesas municipais e detecção de anomalias, com o objetivo de ampliar eficiência dos processos de auditoria das despesas municipais. O principal modelo preditivo apresentado foi elaborado a partir dos dados de despesas de 644 municípios do estado de São Paulo ao longo de 10 anos (2009 - 2019), apresentando um R^2 em torno de 92%.

Tal resultado é promissor para, em conjunto com outros métodos, aumentar a relação custo-retorno dos procedimentos de auditoria das contas públicas e a redução da atratividade de atos de corrupção.

Palavras Chave: Corrupção. Aprendizado de máquina. Detecção de anomalias. Auditorias.

ABSTRACT

The present study, after conducting a literature review of corruption correction methods, and exposing the classic microeconomic models for its analysis, proposes the use of machine learning methods for parameterizing municipal expenses and detecting anomalies, with the aim of to increase the efficiency of the audit processes of municipal expenses. The main predictive model presented was elaborated from the expenditure data of 644 municipalities in the state of São Paulo over 10 years (2009 - 2019), presenting an R^2 of around 92%.

he results are promising. In conjunction with other methods, the use of the model created may increase the cost-return of public accounts audit procedures and reduce corruption among municipalities.

Keywords: Corruption. Machine learning. Anomaly detection. Audit.

SUMÁRIO

1	<i>Introdução.</i>	8
1.1	A corrupção e seus efeitos.	8
1.2	Breves considerações microeconômicas.	9
1.3	A corrupção e sua detecção.	13
1.3.1	Abordagens tradicionais de detecção de corrupção.	15
1.3.2	Utilização de <i>machine learning</i> na detecção e predição de corrupção.	21
1.4	Objetivo e divisão da Dissertação.	24
2	<i>Dados.</i>	26
2.1	Breves considerações sobre os Tribunais de Contas.	26
2.2	Dados das despesas dos municípios do Estado de São Paulo.	28
2.2.1	Tratamento dos atributos selecionados.	36
2.3	Características dos municípios.	38
2.4	Análise Descritiva.	41
3	<i>Modelos e Resultados.</i>	48
3.1	Modelo Linear.	51
3.2	<i>Gradient Booster Regressor.</i>	52
3.3	<i>Random Forest Regressor.</i>	53
3.4	Rede Neural.	54
3.5	Comparação Geral.	54
3.6	Considerações intermediárias acerca da performance em <i>cross validation</i>.	55
3.6.1	Clusterização de municípios.	56
3.6.2	Clusterização por quartil de valor de despesas.	58
3.7	Avaliação no âmbito amostra de teste.	60
3.7.1	Modelo Geral de <i>Random Forest Regressor</i> - Amostra de teste.	61
3.7.2	Modelo <i>Random Forest Regressor</i> por quartil (Amostra de teste).	62
3.8	Avaliação final de performance.	63
3.9	Aplicação prática dos modelos desenvolvidos.	63

4	<i>Conclusão.....</i>	<i>66</i>
4.1	Contextualização dos resultados.	66
4.2	Análise Crítica.....	69
4.3	Pesquisa Futura e Aprimoramentos.....	71
	<i>REFERÊNCIAS.....</i>	<i>73</i>

1 Introdução.

1.1 A corrupção e seus efeitos.

O fenômeno da corrupção é marcado por uma quase ubiquidade, fazendo-se notar em países desenvolvidos ou em desenvolvimento, no setor público ou no privado, em alto ou baixo grau de hierarquia. Não obstante, são nos corredores das repartições públicas de países pobres e em desenvolvimento que as suas sombras se mostram mais frequentes e seus efeitos mais deletérios.

De fato, há evidências significativas de que a corrupção diminui o crescimento econômico. De um lado, a exigência de propina para o regular funcionamento empresarial opera como uma tributação sobre a atividade econômica, mas com alto grau de imprevisibilidade e segredo, gerando efeitos negativos sobre a decisão empresarial, e, assim, afastando o investimento estrangeiro, de acordo com Shleifer e Vishny (1993) e Mauro (1995). Nesse sentido, Fisman e Svensson (2007), analisando os dados de pagamento de propina e de tributação em Uganda, apontam que o impacto marginal do pagamento de propina é três vezes mais negativo sobre o crescimento das empresas do que o de pagamento de tributos.

De outro lado, a corrupção também gera uma perda da capacidade de arrecadação do Estado, possibilitando a evasão de tributos. Adicionalmente, ela afeta o processo de compras públicas, com a perda de recursos e com a inviabilização ou a diminuição da qualidade de projetos de infraestrutura como lembra Olken (2007) e, portanto, deteriorando a formação bruta de capital fixo e aumentando o custo marginal dos investimentos públicos.

Ademais, ponto que por vezes é negligenciado é que, se a exigência de propina tem efeitos quase-tributários sobre o indivíduo, esta extração de renda opera de forma regressiva, com os pobres tendo o seu patrimônio proporcionalmente mais afetado do que os mais ricos. É que, diante de situações de vulnerabilidade, da premência por serviços ou recursos públicos, de restrições a soluções alternativas, os indivíduos mais pobres possuem maior disposição a pagar a propina, sendo mais fortemente afetados, de acordo com Hunt (2006).

Nesse sentido, o modelo microeconômico desenvolvido por Barnerjee, Hanna e Mullainathan (2012) prevê que mecanismos de burocracia excessiva desenhados para gerar discriminação de preços são criados porque o governo fornece bens para indivíduos cuja capacidade de pagar é inferior a sua disposição a pagar, sendo, portanto, um fenômeno mais intenso em governos atendendo aos mais pobres.

A corrupção também distorce as decisões de alocação de recursos públicos, afetando de uma forma mais geral a eficiência e os pontos ótimos das decisões de políticas públicas. Em primeiro lugar, as autoridades públicas que pretendem cometer atos de corrupção tendem a alterar a alocação de recursos para favorecer as categorias de despesas em que a corrupção é de mais difícil detecção, conforme prediz Mauro (1996).

Em segundo, o aumento do custo marginal provocado por corrupção, se antecipado pelo formulador da política pública, pode impedir que um projeto, que seria custo-eficiente, seja implementado, e, portanto, limite a capacidade de desenvolvimento. Por outro lado, se os custos da corrupção não são antecipados, a tomada de decisão governamental deixa de considerar um custo relevante, e, dessa forma, gera o sub-provisionamento de recursos necessários ao atendimento da necessidade da sociedade, nos termos postos por Olken (2012).

Ademais, a corrupção também afeta a capacidade governamental de regular e de induzir o comportamento dos indivíduos. Nesse cenário, a corrupção diminui o impacto de penalidades que são desenhadas para gerar o desincentivo a ações individuais deletérias. Assim, de uma forma mais geral, a capacidade de monitoramento do Estado passa a ser significativamente reduzida, o que facilita a proliferação de atividades com externalidades negativas, como o desmatamento, desordem urbana, poluição e crime organizado. Há, assim, uma deterioração da regulação econômica preconizado por Mauro (1996).

1.2 Breves considerações microeconômicas.

O estudo basilar da aplicação de teoria microeconômica na análise da estrutura de incentivos na decisão de cometimento de atos ilícitos foi o realizado por Gary Becker em 1968 em *Crime and Punishment: An Economic Approach*, e em *paper* derivado deste, em 1973, *Law Enforcement, Malfeasance and Compensation of Enforcers*, em coautoria do próprio e Stigler.

Nesse *framework*, podemos considerar as seguintes abreviações: y é o ganho financeiro decorrente do cometimento de ato ilícito por determinado agente, sendo $y > 0$, p é a probabilidade de detecção de ato ilícito e de aplicação da pena f , sendo $f > 0$. O retorno líquido do cometimento do ato ilícito ($y - f$) e pode ser denominado r . Assim, a utilidade do agente i será $U_i(r)$, com, naturalmente, $\frac{\partial U_i}{\partial r} > 0$.

Assim, a utilidade esperada do agente é:

$$(i) \quad E(U_i) = pU_i(y - f) + (1 - p)U_i(y)$$

Como seria esperado, dessa equação é possível depreender que:

$$(ii) \quad \frac{\partial E(U_i)}{\partial p} = U_i(y - f) - U_i(y) < 0$$

$$(iii) \quad \frac{\partial E(U_i)}{\partial y} > 0$$

$$(iv) \quad \frac{\partial E(U_i)}{\partial f} < 0$$

Esse simples modelo é capaz de gerar intuições relevantes. Em particular, é relevante notar que a utilidade do agente que comete o ilícito é decrescente na probabilidade de detecção do ilícito.

Conforme foi argutamente observado por Zamboni e Litschig (2018), quando trazemos tal modelo ao processo de auditoria, é absolutamente fundamental notar que a probabilidade de interesse não é propriamente a de detecção, mas a de punição efetiva. Nesse sentido, sendo p_a a probabilidade de ser submetido a um procedimento de auditoria condicional a ter cometido um ato ilícito, p_d a probabilidade de detecção condicional a ter sido submetido a um procedimento de auditoria e p_p a probabilidade de sofrer uma punição condicional a ter sido detectada a irregularidade em um procedimento de auditoria, a probabilidade de interesse para a tomada de decisão do agente, é $p = p_a p_d p_p$.

Nesse sentido, há três dimensões em que o processo de auditoria tem potencial para afetar o comportamento do agente que considera o cometimento de um ato ilícito. Pode ser

aumentada a probabilidade de auditoria dos agentes com maior propensão a atos de desvios; pode ser incrementada a probabilidade de serem achados atos ilícitos durante o processo de auditoria; ou, por fim, pode ser ampliada a probabilidade de ser gerada uma punição efetiva caso os atos ilícitos tenham sido detectados.

Cada um desses aspectos demanda abordagens distintas de aprimoramento do processo de auditoria. Nesse sentido, como veremos, métodos classificatórios de ciência de dados têm sido utilizados para apontar setores ou entidades mais propensas a risco de cometimento de ilícitos, aumentando p_d .

Outra abordagem, a que motiva os modelos construídos nesta Dissertação, é a criação de métodos de detecção de anomalias, de modo amentar p_d . Ainda, é relevante notar que a relevância de p_p implica que métodos que busquem o aumento da potencialidade das atividades de monitoramento também devem cuidar para que os achados gerados sejam úteis a informar um processo punitivo. Nesse sentido, é relevante que sejam explicáveis e gerem evidências suficientes a serem sustentáveis diante de contestação jurídica.

É ainda relevante notar que o planejador público escolherá como ponto ótimo de nível de auditoria aquele em que o benefício marginal de aumento de p equipare-se ao custo marginal do processo de auditoria. Nesse sentido, mecanismos que impactem este último também trarão resultados positivos ao desenho de mecanismos para a diminuição de atos ilícitos.

A teoria desenvolvida a partir do trabalho de Becker abarca outros aspectos da dinâmica dos atos ilícitos. Um que gerou diversas revisões e debates é a teoria de salário eficiência, em que Becker argumenta que uma estrutura de incentivos possível é o oferecimento de salários acima de mercado, de modo que o custo de uma eventual detecção e demissão seja trazido a valor presente pelo agente, dissuadindo-o ao cometimento de irregularidade.

Se o modelo de salário eficiente de fato se verificar empiricamente¹, deveria ser avaliado o *value for money* de diferentes estratégias de incentivos para a diminuição de ilícitos, comparando-se os custos do processo de auditoria, em uma estratégia punitiva, ou do aumento de salários dos agentes. De toda forma, quanto maior a probabilidade de detecção, menor precisaria ser a diferença de salários em relação ao mercado, sendo, assim, abordagens complementares.

Trocando-se os sinais do salário eficiência, também é possível trazer intuições relevantes para a análise da corrupção enquanto fenômeno econômico. Nesse sentido, a eventual verificação de salários de servidores públicos abaixo daquele que seria disponível em mercado, controlando para fatores de diferenciação, como estabilidade e nível de qualificação, pode ser um indício de que os retornos esperados são compensados por atos de corrupção ou de diminuição de esforços.

Outras abordagens microeconômicas do ilícito foram desenvolvidas para a avaliação da dinâmica da corrupção. Nesse sentido, o estudo já clássico de Shleifer e Vishny (1993) avalia a relevância de levar em consideração os mecanismos de organização da atividade dos agentes públicos na determinação da capacidade destes de extrair renda por atos de corrupção. Nesse sentido, os autores avaliam o impacto de desenhos institucionais que abordam relações estratégicas, considerando elementos de competição, competição monopolística e de dupla marginalização para a determinação do preço da propina, e suas consequências no desenho de políticas públicas.

Outro estudo de significativo impacto foi o *Working Paper* de Banerjee Et Al (2012). Neste, os autores constituem elaborada modelagem microeconômica para considerar as relações entre principal e agente em um contexto de assimetria de informações. Assim, é desenhado um jogo em que o agente deve decidir pela alocação de recursos públicos seguindo ou não a regras impostas pelo planejador público.

¹ Alguns *papers* recentes buscaram evidências empíricas da teoria do salário eficiência. Em estudo relevante, Di Tella e Schargrodsky (2003) verificaram evidências de confirmação do impacto dos salários dos servidores na diminuição de atos de corrupção no âmbito de compras públicas de hospitais em Buenos Aires. De uma forma geral, a objeção teórica ao argumento do salário eficiência centra-se na argumentação de que os potenciais ganhos com atos de corrupção seriam excessivamente altos para serem realisticamente compensados por expectativa de diferencial de salário em relação ao mercado.

A decisão do agente é exclusivamente pautada pela maximização de sua utilidade, considerando as restrições de demanda do mercado em que atua, da probabilidade de detecção de quebra da regra e do nível da respectiva punição. Por outro lado, o planejador público deve decidir pelo desenho da regra, conferindo maior ou menor grau de discricionariedade, considerando o comportamento esperado do agente, que, inclui a possibilidade de corrupção.

O desenho de regras que concedem maior grau de discricionariedade ao agente pode, em determinadas circunstâncias, ser a decisão *first best* do planejador. No entanto, diante de assimetria de informações, o controle da decisão do agente passa a ser imperfeito, o que pode gerar a necessidade de regras mais restritas em uma decisão *second best*.

Ao fim, a análise da microeconomia da corrupção é um campo ativo estudo e de impacto significativo no desenvolvimento de políticas públicas. Na medida em que o objeto dessa dissertação se insere no campo da detecção de atos lesivos à administração pública, o desenvolvimento extensivo de cada uma das abordagens teóricas microeconômicas foge ao escopo pretendido. Não obstante, é relevante ter em consideração tais construções, a uma, porque podem trazer relevantes intuições para a identificação de elementos preditivos do ato ilícito, a duas, porque ajudam a entender de que forma o monitoramento pode gerar benefício para a sociedade.

1.3 A corrupção e sua detecção.

Até agora, tratamos do termo corrupção de forma excessivamente intuitiva, sem defini-la propriamente. Adotaremos, ao longo desta dissertação, uma definição muito próxima à proposta por Barnerjee Et Al (2012): a violação de uma regra por um agente público para auferir benefícios para si ou para outrem.

É de se notar que a forma como trataremos corrupção não se vincula às tipologias postas pela legislação brasileira ou de outro país em específico. A razão desse descompasso é que o objetivo deste trabalho é muitos mais servir a desenhos de mecanismos institucionais do que a atividades jurisdicionais. Dentro de nossa definição, cabem diferentes tipos penais: corrupção ativa, corrupção passiva, prevaricação, concussão, falsidade ideológica, peculato,

dentre diversos outros, ou mesmo ilícitos administrativos como a improbidade administrativa. A tipificação em si caberá ao operador do direito.

O que se pretende abarcar na definição utilizada é o típico problema de agência em um cenário de informação assimétrica, no âmbito da implementação de políticas públicas. É relevante marcar que, ainda que um formulador de políticas públicas busque a maximização do bem estar social, na medida da impossibilidade de um monitoramento perfeito, este se vê diante da possibilidade sempre presente de que a regra criada não seja executada pelo agente público, quando a violação desta regra representar a maximização da utilidade do agente. É este grupo de ações que se pretende abarcar. Nesse sentido, enquadram-se na definição trazida ações claramente percebidas dentro do conceito de corrupção, mas também atos nem sempre entendidos como tal, como por exemplo faltas recorrentes ao trabalho e a adoção irregular de jornadas reduzidas.

Ocorre que, pela própria natureza da corrupção, esta é secreta, de difícil apreensão objetiva, o que gera o enorme desafio para a sua detecção. Atenta à relevância do tema, nas duas últimas décadas, a academia foi prolífica no desenvolvimento de métodos para ampliar a capacidade de detecção ou mensuração de atos de corrupção.

Nesta subseção, passamos a expor os resultados de pesquisa bibliográfica acerca desses métodos. Iniciamos com a apresentação de abordagens tradicionais, que, em geral, possuem um enfoque de uso acadêmico, seguindo em grande parte as referências e categorias trazidas pelo relevante *paper* de Olken (2011). Nessa etapa, são apresentadas as seguintes abordagens:

- a) Mensuração por percepção;
- b) Mensuração por observação direta e pesquisa com envolvidos;
- c) Estimação por diferenças de registros;
- d) Estimação por equilíbrio de mercado;
- e) Avaliação por resultados de auditoria.

Com a exceção, do uso de avaliações de auditoria, todos os outros métodos possuem pouca aplicação enquanto instrumento de monitoramento contínuo dos atos dos agentes

públicos e seriam de difícil aplicação, portanto, para a dissuasão do comportamento lesivo à administração pública que é o fim último desta Dissertação. Não obstante, a sua exposição serve ao propósito de contextualização e progressão de métodos nesta seara, dando um dimensionamento do problema que se coloca.

Na subseção subsequente, são expostos os métodos desenvolvidos de análise de corrupção por meio da aplicação de algoritmos de *machine learning*. Tais métodos possuem o mérito de gerar uma potencial aplicação prática e contínua nos procedimentos de controle e, em nosso entendimento, ensejam um relevante campo de pesquisa a ser desenvolvido. Esta Dissertação insere-se exatamente com o objetivo de tentar trazer contribuições relevantes a essa nova abordagem.

1.3.1 Abordagens tradicionais de detecção de corrupção.

A) Mensuração por percepção.

O método que iniciou um amplo grupo de pesquisas por Mauro (1995) na avaliação de causas e efeitos de corrupção foi a utilização de rankings qualitativos de percepção do nível de corrupção em diferentes países. Tais análises são realizadas por quadros de referência como o *Corruption Perception Index* (CPI) da Transparência Internacional² e o *Control of Corruption Index* do Banco Mundial³. Mais recentemente, provedores de análise de risco político também passaram a prestar análises qualitativas do nível de corrupção de países como um dos elementos para a caracterização dos índices de risco político⁴.

A despeito da grande relevância dessas análises para os estudos de corrupção, tais métodos possuem significativas restrições. Em primeiro lugar, eles carregam um alto grau de subjetividade. Nesse sentido, como notou Olken (2007), os indivíduos têm maior capacidade para captar determinados aspectos da corrupção do que outros. Além disso, tais índices capturam o que se percebe de corrupção, e não a corrupção efetivamente. Assim, distinções no

² <https://www.transparency.org/en/cpi/2019/index/nzl>

³ https://tcdata360.worldbank.org/indicators/hc153e067?country=BRA&indicator=364&viz=line_chart&years=1996,2019

⁴ <https://www.prgroup.com/explore-our-products/international-country-risk-guide/>

grau de transparência, liberdade de imprensa, ou da robustez das instituições, afetam a percepção, ainda que não haja diferenças no nível de corrupção.

Ademais, o escopo de análise dos índices de percepção de corrupção usualmente disponíveis é limitado, sendo excessivamente gerais, o que restringe a sua utilização para fins de formulação de políticas e desenhos institucionais.

B) Observação direta e pesquisas com envolvidos.

Uma outra forma⁵ de se tentar mensurar a corrupção é por meio de análises de evidências diretas de corrupção por vezes encontradas em investigações ou por meio de colaboração direta dos envolvidos. Ocorre que o acesso a esse tipo de informação é raro, e, em que pese possam servir para o teste de hipóteses teóricas do comportamento dos agentes, é de relevância limitada para que passa ser utilizada como um instrumento contínuo de monitoramento.

De outra parte, há estudos⁶ que tentam realizar pesquisas com potenciais participantes ou vítimas de corrupção. No entanto, na medida em que se trata do cometimento de ilícitos, tal abordagem é excessivamente restrita, tendendo a capturar, ao máximo, a declaração do cometimento de ilícitos menores.

C) Estimação por diferenças de registros.

Um grupo de métodos apresentado por Olken (2007) que trouxe avanços significativos para a detecção e monitoramento dos processos de corrupção é o que implementa avaliação de diferenças entre registros distintos de quantidades do mesmo elemento. Olken (2007) aponta como estudo pioneiro nesse sentido o realizado por Reinikka and Svensson (2004), no qual os autores apuram as diferenças entre os registros de recursos transferidos para as escolas de Uganda com os efetivamente recebidos pelos alunos.

⁵ OLKEN, Benjamin A.; e PANDE, Rohini. **Corruption in Developing Countries». Annual Review of Economics 4**, n°. 1. Setembro de 2012.

⁶ <https://www.enterprisesurveys.org/>; <https://wp.unil.ch/icvs/>

Ainda, Olken (2007) relata o trabalho de Fisman and Wei (2004) que compara as diferenças entre o registrado como exportado e o importado nas relações comerciais Índia-China e verifica as diferenças entre os valores de custos de rodovias registrados e os efetivamente realizados.

Os métodos de diferenças em registros representam um grande avanço no sentido de se construir meios para aferir objetivamente alguns resultados de corrupção. Além disso, trata-se de um método relativamente universalizável, podendo ser aplicado sistematicamente em diferentes esferas, e pode servir, inclusive para a formulação de políticas no combate a corrupção.

Por outro lado, é preciso marcar que tais métodos também possuem limitações relevantes. Em primeiro lugar, o que é um problema quase inafastável nesta seara, as diferenças encontradas entre os dois pontos de quantificação não necessariamente derivam de atos de corrupção, mas podem simplesmente representar perdas naturais, ou erros de registro. Mais relevante, contudo, é que tais métodos pressupõe a existência ao menos de dois pontos diferentes confiáveis de mensuração da mesma variável de interesse. No entanto, de uma forma geral, tais pontos independentes de avaliação não estão presentes.

De outra parte, tais métodos apenas são relevantes para verificar atos de corrupção realizados por subtração de recursos. No entanto, este é apenas um subconjunto relevante de atos de corrupção, deixando de fora, por exemplo, o recebimento de propina pela escolha da alocação de recursos escassos, por exemplo. Nesses casos, não haverá diferença de registros nas contas públicas.

Não obstante as ressalvas apontadas, tais métodos são de significativa relevância no aprimoramento de técnicas para a detecção de atos de corrupção, e, em grande medida são incorporados nas técnicas contemporâneas.

D) Estimação por equilíbrio de mercado.

Uma abordagem relevante para a detecção de mercado é a utilização de conceitos como de informação perfeita no mercado de ações ou de condições de equilíbrio no mercado de trabalho.

No âmbito da primeira abordagem, são analisadas como que o valor de mercado de empresas listadas em bolsa e consideradas conectadas com figuras políticas relevantes é afetado por eventos relacionados a essas pessoas, como a sua morte, deposição ou perda de poder político⁷.

Já pelo segundo método, são avaliadas diferenças no salário entre agentes públicos e de empregados no setor privado com qualificações equiparáveis. Se, mesmo após utilizar métodos para controles de elementos diferenciadores como estabilidade de emprego, é verificado que os agentes públicos são sub-remunerados, é razoável inferir que os agentes públicos conseguem extrair, pela sua posição, rendas adicionais que compensam a diferença de remuneração, seja por subtração de recursos públicos, recebimento de propina, ou simplesmente por trabalhar um número de horas aquém do contratado.

E) Avaliação por resultado de auditorias.

A utilização dos relatórios de auditoria é um instrumento de significativo impacto na prevenção a atos de corrupção. Seguindo a construção teórica elaborada por Ferraz e Finan (2016), o processo de auditoria de contas públicas afeta os agentes políticos sob dois diferentes mecanismos: político e jurídico-punitivo.

Sob o impacto político, a transparência de resultados negativos de auditoria gera efeitos no processo de reeleição, sendo um elemento de seleção que reduz a corrupção futura, ao mesmo tempo que insere o custo do impacto no processo eleitoral na utilidade dos agentes políticos, reduzindo a propensão ao cometimento de ilícitos. Por outro lado, conforme já

⁷ FISMAN, Raymond; e GATTI, Roberta. **Decentralization and Corruption: Evidence Across Countries?** Sem data.

desenvolvido, os processos de auditoria aumentam a probabilidade de detecção e punição jurisdicional, o que incrementa o custo dos atos de corrupção pelos agentes públicos.

Nesse sentido, Ferraz e Finan utilizaram os dados disponíveis das auditorias aleatórias realizadas pela Controladoria Geral da União (CGU) sobre a utilização de recursos federais pelos municípios brasileiros para demonstrar que, entre aqueles que foram submetidos a procedimentos de auditoria, há uma redução de 8% no nível de corrupção em relação a municípios não auditados. Além disso, é verificado um efeito de *spillover*, com municípios vizinhos que foram submetidos a auditoria também apresentando reduções no nível de corrupção da ordem de 7.5%.

É ainda relevante destacar que Ferraz e Finan tentam diferenciar o impacto entre os efeitos políticos e jurídico-punitivo e encontram o relevante resultado de que o último é responsável por 94% da redução encontrada de corrupção. Tal evidência é fundamental no desenho de políticas para o combate à corrupção, mostrando a relevância de mecanismos que aumentam a intensidade da punição esperada do agente.

Zamboni e Litschig (2013), também trabalhando com os dados das auditorias aleatorizadas da CGU, estimam que o aumento da probabilidade de ser submetido a um procedimento de auditoria reduz em 17% os apontamentos relacionados a má-gestão ou corrupção no âmbito das compras públicas. Adicionalmente, tal aumento de probabilidade também está associado a uma modificação das modalidades de compras adotadas pelos gestores públicos em direção a métodos que diminuem os graus de liberdade do gestor.

Ao mesmo tempo, os autores não verificaram relação significativa entre o aumento da probabilidade de auditoria e a qualidade dos serviços públicos. Como possível causa dessa diferença de resultado, o estudo aponta as distinções entre estruturas punitivas (sendo mais intensa no primeiro caso), e a dificuldade de detecção de irregularidades (sendo mais difícil de ser verificada no segundo). Uma possível explicação adicional, a ser verificada empiricamente, é se a diferença de resultado também não pode ser consequência do horizonte temporal de análise. Ao passo que os efeitos das alterações de compras são imediatos, é possível que, para que sejam verificados efeitos na qualidade de serviço, a modificação do comportamento dos agentes públicos deva ser contínua e de longo prazo.

Di Tella e Schargrodsky (2003) realizaram importante avaliação da relação entre o processo de auditoria e salários dos agentes públicos em um forte movimento de combate a corrupção no âmbito de compras públicas dos hospitais em Buenos Aires. O trabalho aponta que há impacto de salários mais altos de agentes públicos na diminuição de corrupção em cenários de intensidade de auditorias moderadas, não tendo efeitos significativos diante de auditorias muito baixas ou muito altas. Por outro lado, o estudo aponta a dificuldade de se manter níveis intensos de auditoria efetivo em um longo período, de modo que a combinação de estruturas de auditoria/punitiva e salário eficiência é um instrumento relevante para a diminuição da corrupção dos agentes públicos, confirmando o modelo de Becker e Stigler (1974).

Ainda, Olken (2007) realiza a avaliação de procedimentos de auditoria em pequenos projetos de manutenção de estradas na Indonésia. Neste estudo, há a avaliação de impacto por aleatorização, em que é demonstrado que vilas que eram comunicadas que seriam submetidas a procedimentos de auditoria previamente ao início das obras, saindo de uma probabilidade base de 4% para de 100% de auditoria, apresentavam uma redução de 8% do nível de perdas de recursos transferidos.

Outro aspecto de grande relevância trazido pelo estudo de Olken (2007) é a demonstração das diferenças entre a auditoria governamental e o controle pela comunidade local afetada. É verificado que, ao menos para o caso avaliado, a ação da comunidade local era efetiva para expor os casos em que havia o impacto individual direto, como na remuneração dos trabalhadores. No entanto, para os casos de impacto geral na comunidade o problema de *free riding* aparece, não sendo demonstrado impacto relevante da ação de controle comunitários. Ao contrário, as ações de auditoria eram efetivas em ambos os casos. Tal análise é relevante na escolha de modelos de controle, incluindo as diferenças de custo-retorno.

Outro aspecto fundamental a marcar no estudo é que, a despeito da antecipação do processo de auditoria e, muito embora tenha sido verificada uma diminuição de 8% das perdas não explicadas de recursos, ainda houve uma perda de 19.7% de recursos acima do esperado. O autor relata que os procedimentos de auditoria efetivamente verificavam irregularidades, no entanto, não eram capazes de apontar peremptoriamente evidência de cometimento de ilícitos

que suscitassem provável punição. É absolutamente necessário frisar, portanto, o desafio de se construir métodos que aumentem o grau de eficiência dos procedimentos de auditoria.

1.3.2 Utilização de *machine learning* na detecção e predição de corrupção.

Uma nova abordagem para o tratamento de corrupção é a utilização de técnicas de *machine learning* como instrumentos preditivos que aumentem a eficiência de procedimentos de auditoria e auxiliam no desenho de mecanismos preventivos à corrupção. Conforme aponta Gallego Et Al. (2018), embora a maior parte dos estudos passados sejam focados em avaliar a causalidade dos atos de corrupção e tenham tido imensa importância para a compreensão deste fenômeno, de uma perspectiva de desenho de políticas públicas, a realização de predições talvez seja ainda mais relevante.

De fato, o desenvolvimento de instrumentos que sejam capazes de processar dados e realizar um *screening* de atos prováveis de corrupção é fundamental a pelo menos quatro aspectos: (i) aumento da probabilidade de submissão de atos de corrupção à avaliação de auditoria; (ii) incremento da qualidade da análise das auditorias realizadas; (iii) melhoria da relação custo-retorno dos procedimentos de controle, e (iv) melhoria dos desenhos de mecanismos.

O problema de agência que sustenta atos de corrupção deriva da incapacidade de realização de um monitoramento perfeito das atividades dos agentes públicos, em função da existência de custo das atividades de controle. Nesse caso, a Administração deve observar a relação custo retorno da implementação de monitoramento.

Neste cenário, o planejador precisa determinar o que e quem será auditado. Diante da ausência de métodos robustos de predição de atos ilícitos, a administração não é capaz de realização da discriminação de tipos de agentes ou atividades, impedindo a distribuição ótima dos esforços de controle (afasta-se do desenho *first-best*). Como solução *second-best*, em geral, adota-se uma aleatorização ou semi-aleatorização das escolhas de objeto de auditoria.

Assim, o agente público admite uma probabilidade de detecção de seu ato ilícito, passando a ter a sua utilidade como uma loteria. Sua punição depende da probabilidade de ser

submetido à auditoria, e da probabilidade de punição condicional a ter sido submetido a esse procedimento.

Um modelo preditivo pode gerar *scores* classificatórios que indiquem determinada entrada de análise como provavelmente relacionada a ato de corrupção. Por meio desse instrumento, a Administração passa a realizar a discriminação de tipos de agentes públicos ou de atividades em termos de probabilidade de cometimento de atos ilícitos. Assim, atividades de controle podem alterar a distribuição de alocação de esforços de auditoria, concentrando entre aqueles com maior probabilidade de cometimento de atos ilícitos, aproximando-se de um resultado *first-best*. Tal mudança tende a alterar a percepção dos agentes propensos ao cometimento de atos ilícitos de sua probabilidade de ser submetido a auditoria, aumentando a expectativa de punição, e diminuindo o resultado esperado da corrupção.

Recentes estudos foram desenvolvidos nesse propósito. Gallego (2018) utiliza a detalhada base de contratos públicos das municipalidades da Colômbia para cruzar com apontamentos de irregularidades realizados pelo sistema jurisdicional colombiano e desenvolver um sistema classificatório pulverizado, em que são oferecidos *scores* preditivos de que determinado contrato é propenso a ser objeto de ilícito. Para a realização desse sistema classificatório preditivo, Gallego utiliza algoritmos supervisionados de *machine learning*: *lasso regression*, *gradient boosting machine* e *classification trees*, obtendo em todos os casos AUC⁸ acima de 80%, analisando mais de 2 milhões de contratos.

De forma semelhante, Grace Et Al. (2016) desenvolveram sistema automatizado de detecção de irregularidades no âmbito de contratos do Banco Mundial em países em desenvolvimento. Os autores elaboraram sistema classificatório supervisionado com base na relação de contratos do banco e apontamentos de irregularidades registrados no sistema interno do Banco Mundial derivado de histórico de investigações iniciadas por denúncias de irregularidades. Após testar os algoritmos *random forest classifier*, *logistic regression*, *ada boost classifier*, *support vector machine*, *k neighbors classifier*, e *gradient boosting machine*,

⁸ AUC (área abaixo da curva ROC) é um comum medidor de performance de modelos classificatórios de *machine learning*. A ROC (*receiver operating characteristic curve*) é uma curva de performance que mostra a relação entre a taxa de falsos positivo e positivos verdadeiros para cada *thresholds*.

este último obteve a melhor performance, com uma precisão próxima a 80% entre os 25% cuja probabilidade de ilicitude foi apontada pelo modelo como maior.

Em abordagem um pouco distinta, Ash Et Al (2020) utilizam os dados de finanças públicas e outras características de municípios brasileiros e os dados de apontamentos de irregularidades das auditorias aleatorizadas da Controladoria Geral da União (CGU) que avaliam a aplicação de recursos derivados de transferências federais para elaborar, com o uso de *gradient booster classifier*, um modelo classificatório dos municípios em relação à existência de atos de corrupção, com uma acurácia da predição de 76%.

Os autores ainda realizam uma avaliação de importância das variáveis preditivas e verificam uma correspondência entre os elementos considerados pelo modelo preditivo e os fundamentos apresentados nos relatórios de auditoria do órgão de controle. Outro aspecto relevante é que, a partir do modelo criado, os autores são capazes de gerar dados sintéticos de corrupção e com eles estimar o impacto das auditorias. Na análise realizada, é confirmado o impacto significativo das auditorias na redução de corrupção não só dos municípios avaliados como em municípios vizinhos, confirmando o efeito de *spillover* já apresentado por Ferraz e Finan (2016).

Collonelli Et Al (2020), também trabalhando com dados de municípios brasileiros e relatórios de auditoria da CGU, implementam modelos preditivos classificatórios de corrupção para avaliar quais elementos fornecem maior valor preditivo. Os autores analisam os algoritmos *random forest classifier*, *gradient booster classifier*, *lasso*, *super learner ensemble*, e redes neurais, verificando uma maior performance entre os dois primeiros. No âmbito na análise de *feature importance*, os autores apontam que os elementos que mais fortemente predizem elementos de corrupção no âmbito da utilização pelos municípios das transferências federais, surpreendentemente, são os relacionados à organização local do setor privado, seguido de características de desenvolvimentos financeiro, demografia e capital humano.

Por fim, um recente artigo de Gallego Et Al (2020) é exemplificativo das possibilidades geradas pela utilização dos modelos preditivos aqui listados. Diante da crise posta pela pandemia do COVID-19, os autores utilizam modelo de classificação de corrupção no âmbito dos contratos das municipalidades da Colômbia, para verificar de que forma as

medidas adotadas de flexibilização das regras de compras públicas realizadas em função da crise se relacionavam com a classificação gerada. Assim, os autores verificam que os municípios classificados como mais propensos a corrupção são os que reagem às modificações regulatórias com a adoção de mecanismos mais discricionários de compras públicas.

Todos esses recentes estudos demonstram a enorme potencialidade em termos de mecanismos de combate e análise de corrupção que os modelos preditivos de *machine learning* geram. Esta Dissertação insere-se neste contexto. Apresentaremos nos capítulos que seguem a formulação, com uso de algoritmos de *machine learning*, de modelo preditivo de gastos dos municípios do Estado de São Paulo, a partir da base de dados disponibilizadas pelo Tribunal de Contas de São Paulo – TCE/SP e de dados de características municipais tomadas da base de dados do IBGE. O objetivo desta abordagem é que o modelo preditivo possa servir de parâmetro de níveis esperados das despesas. Assim, esperamos contribuir para o aumento do retorno marginal dos esforços de auditoria, diminuindo os custos necessários à identificação de anomalias nas despesas municipais.

1.4 Objetivo e divisão da Dissertação.

A presente dissertação é dividida em 4 seções. Em primeiro plano, nesta Introdução, apresentamos a relevância do tema, trouxemos breves referências de modelos microeconômicos de análise de corrupção e de seu monitoramento, realizamos uma revisão bibliográfica acerca de mecanismos de detecção da corrupção e uma exposição de como o modelo que propomos de parametrização e de detecção de anomalias de despesas públicas de municípios do Estado de São Paulo insere-se em tal contexto.

Na segunda seção, realizamos a análise descritiva dos dados utilizados para o desenvolvimento dos modelos preditivos. Nesta, mostramos de que forma manipulamos os 120 GB de dados de despesas fornecidos pelo TCE-SP por meio da *sqlite*, fundamentamos a seleção de dados desta base, assim como as opções de agregação dos dados. É também nesta etapa que apresentamos o impacto de restrições computacionais nas opções de modelagem. Por fim, provemos referências estatísticas e gráficas das distribuições de dados, mostrando os seus impactos nos modelos criados.

Na terceira seção, passamos a expor modelos de uso de algoritmos de *machine learning* desenvolvidos e testados para a predição das despesas dos municípios de São Paulo, assim como os seus resultados. Apresentamos, de início, alguns resultados de referência como a mera adoção da média ou da mediana como parâmetro preditivo. Em seguida, ainda com um objetivo de fornecer referência de performance e também de verificação de linearidade, mostramos os resultados de um modelo preditivo linear simples.

Ainda na mesma etapa, uma vez expostos resultados de referência, apresentamos os modelos efetivamente desenvolvidos e testados, quais sejam, *Random Forest*, *Gradient Booster Regressor* e redes neurais. Todos os resultados são apresentados em *cross validation* em cinco *folds*. Mostramos que o modelo de melhor performance foi o de *Random Forest*. A partir deste, testamos os seus resultados em amostra de teste segregada de treino e avaliamos a dispersão de erros. Adicionalmente, em função destes resultados, propomos abordagens alternativas com vista a mitigar o impacto do significativo desbalanceamento dos dados, realizando dois tipos de segregações preditivas: por *cluster* de municípios, e por quartil de nível de despesa. Ainda, mostramos uma análise do uso prático do modelo pelo TCE/SP.

Por fim, na seção de conclusão, fazemos apontamentos finais, relacionando o modelo preditivo desenvolvido com os aspectos teóricos apresentados nesta Introdução. Faremos uma avaliação crítica dos resultados, apresentando as suas possíveis contribuições a esse rico e recente processo de aplicação de mecanismos de *machine learning* na detecção de atos lesivos à Administração Pública. Também apontaremos restrições e fragilidades do *framework* construído e indicaremos possíveis caminhos de aprimoramento em pesquisas futuras.

2 Dados.

Nesta seção passaremos a expor a origem do trabalho empírico que fundamenta esta dissertação, qual seja, a parceria desenvolvida entre a Fundação Getúlio Vargas e o Tribunal de Contas do Estado de São Paulo. Além disso, teceremos breves considerações acerca deste órgão de controle e seus procedimentos de auditoria, e, por fim, realizaremos a análise descritiva dos dados que foram utilizados para a modelagem do nosso modelo preditivo.

A Fundação Getulio Vargas celebrou parceria com o Tribunal de Contas de São Paulo – TCE/SP com o objetivo de desenvolver métodos estatístico-computacionais para a parametrização dos gastos de 644 municípios do estado de São Paulo. Esta dissertação insere-se dentro desse propósito.

O TCE/SP disponibilizou dados elaborados pelo Auditoria Eletrônica de São Paulo - AUDESP, que constitui um sistema de registro e contabilização das despesas de todos os 644 municípios do Estado de São Paulo, com a exceção da capital. Tais dados, que são divulgados publicamente pelo TCE/SP⁹, constituem um importante elemento de transparência e orienta em grande parte o processo de auditoria.

Para fins da modelagem desenvolvida, adicionamos a estes dados que buscam capturar características dos Municípios para que sirvam de variáveis de controle das diferenças das despesas municipais. Esta última categoria de dados foi obtida ora pela base de dados do IBGE¹⁰, ora por dados fornecidos diretamente pelo TCE/SP.

2.1 Breves considerações sobre os Tribunais de Contas.

O Tribunal de Contas é um órgão governamental autônomo, com previsão constitucional, responsável pela análise dos gastos públicos, cuja ação fiscalizadora, denominada controle externo, é auxiliar (mas não subordinada) àquela exercida pelo Poder Legislativo. Nesse sentido, cabe ao Poder Legislativo, que compreende o Senado Federal, a

⁹ <https://transparencia.tce.sp.gov.br/conjunto-de-dados>

¹⁰ <https://sidra.ibge.gov.br/tabela/5938>

Câmara dos Deputados, as Assembleias Legislativas e as Câmaras de Vereadores, a titularidade deste controle externo, frente aos representantes do Poder Executivo, que por sua vez estão representados pelo Governo Federal, Governos Estaduais e Prefeituras dos Municípios.

Os diversos Tribunais de Contas possuem como principal função a fiscalização contábil, financeira, orçamentária, operacional e patrimonial dos órgãos da Administração Pública direta e indireta.

Tratando-se de tribunais administrativos, também julgam as contas dos administradores públicos e demais responsáveis pelo recebimento de verbas ou administração do patrimônio público, ou seja, fiscalizam os órgãos governamentais dentro de sua esfera de atuação, além das contas de qualquer ente ou pessoa a eles relacionada, cujas ações possam resultar em prejuízos ao erário. Nesse contexto, verificadas irregularidades na aplicação do dinheiro público, os Tribunais Contas também podem aplicar sanções aos responsáveis, como multas pecuniárias e proibição de contratar com a administração pública.

No exercício de sua função fiscalizadora, são utilizadas como instrumentos as auditorias, que podem ser realizadas por iniciativa do próprio Tribunal ou por requerimento do Poder Legislativo, para apuração de denúncias, para apreciação da legalidade de atos de concessão de aposentadorias, reformas, pensões, admissão de pessoal no serviço público, fiscalização de renúncia de receitas, além de atos e contratos administrativos em geral.

Na área governamental, há duas formas de auditoria, quais sejam, de conformidade/regularidade e operacional. O Tribunal de Contas da União – TCU em seu glossário, disponível na sua página da internet, conceitua auditoria de conformidade como:

procedimento de fiscalização utilizado para obter dados de natureza contábil, financeira, orçamentária e patrimonial, quanto aos aspectos técnicos, de legalidade e de legitimidade da gestão dos responsáveis pelo órgão, projeto, programa ou atividade e para conhecer a organização e o funcionamento dos órgãos e entidades da Administração direta, indireta e fundacional dos Poderes da União, inclusive fundos e demais instituições jurisdicionadas.¹¹

¹¹<https://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?fileId=8A8182A25EC0F015F58F31F6504A6>

Já a auditoria operacional, é aquela que pretende avaliar o desempenho e a eficácia das operações, além da adequação e oportunidade das políticas públicas e decisões estratégicas da administração pública. Ainda segundo glossário da Corte de Contas Federal, auditoria operacional é a:

avaliação independente, objetiva e confiável que analisa se empreendimentos, sistemas, operações, programas, atividades ou organizações do governo estão funcionando de acordo com os princípios da economicidade, eficiência e efetividade e se há espaço para aperfeiçoamento (ISSAI 300). Exame independente e objetivo da economicidade, eficiência, eficácia e efetividade de organizações, programas e atividades governamentais, com a finalidade de promover o aperfeiçoamento da gestão pública (TCU, 2010).¹²

Nesse contexto, utilizamos a base de dados do Tribunal de Contas do Estado de São Paulo (TCE-SP), que foi pioneiro com a criação da Auditoria Eletrônica de São Paulo (AUDESP), entre 2012 e 2015, com o escopo de fazer o processamento eletrônico de dados de órgãos públicos municipais e estaduais.

Dessa forma, é disponibilizado para as partes inseridas, no processo de fiscalização, um banco de dados de cada órgão, sobretudo com seus dados contábeis, alimentado pelos próprios órgãos a serem fiscalizados. As referidas informações desse banco de dados são ainda fornecidas, pela Internet, à sociedade em geral, o que também contribui de forma significativa para o aprimoramento da transparência dos gastos públicos.

2.2 Dados das despesas dos municípios do Estado de São Paulo.

¹²<https://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?fileId=8A8182A25EC0F015F58F31F6504A6>

O Sistema AUDESP constitui base de dados detalhada das despesas dos municípios paulistas. Ao todo, são disponibilizadas ao público as seguintes características de despesas:

Tabela 1 – Atributos das despesas dos municípios de São Paulo

Atributo	Descrição
id_despesa	Identificação única de despesa
ano_exercicio	Ano do exercício financeiro da despesa
ds_municipio	Nome do município responsável pela despesa.
ds_orgao	Nome do órgão municipal responsável pela despesa.
mes_referencia	Identificação numérica do mês de realização da despesa
mes_referencia_extenso	Mês de realização da despesa por extenso
tp_despesa	Estágio da execução orçamentária (empenhada, liquidada, pago, anulado e reforço)
nr_empenho	Número de registro do empenho
identificador_despesa	Documento de identificação da contraparte
ds_despesa	Nome de identificação de contraparte
dt_emissao_despesa	Data de emissão da despesa – Formato d/m/a
vl_despesa	Valor das despesas em reais
ds_funcao_governo	Categoria, a partir de classificação fechada, da função geral de governo que fundamenta a despesa.
ds_subfuncao_governo	Sub categoria de ds_funcao_governo, também delimitada por lista fechada.
cd_programa	Código de programa de governo a que a despesa está vinculada.
ds_programa	Nome do programa de governo a que a despesa está vinculada.
cd_acao	Código da ação de governo a que a despesa se vincula.
ds_acao	Nome da ação de governo que fundamenta a despesa.
ds_fonte_recurso	Descrição da fonte de recursos para o pagamento da despesa.
ds_cd_aplicacao_fixo	Descrição, com id de identificação da relação da categoria geral da despesa (ds_funcao_governo) com o tipo de fonte de recursos.
ds_modalidade_lic	Modalidade de licitação, sua dispensa, ou inexigibilidade no processo de seleção de contraparte e dos produtos ou serviços adquiridos pela despesa.
ds_elemento	Descrição da despesa em categorias mais granulares que conta com sistema de código

	de identificação que vincula às categorias mais amplas acima descritas.
historico_despesa	Campo livre para a inserção de informações adicionais relevantes.

Fonte: Elaboração própria.

Nesse passo, é necessário trazer algumas considerações de ordem técnica e prática que nos levaram a escolher determinados atributos de despesas como variáveis de controle em detrimento de outras.

Em primeiro lugar, ao todo, temos disponíveis dados individuais de despesas de 644 municípios entre os anos de 2010-2019, que somam em torno de 120 GB de dados. Dadas as limitações de processamento computacional e de memória, de partida, era claro que teríamos uma inviabilidade de gerar modelos de *machine learning* a partir dos dados individualizados de despesas. Sendo assim, optamos por trabalhar com a agregação das despesas de forma anual, agrupando-as pelas seguintes características: Município (*ds_municipio*), órgão municipal que realizou a despesa (*ds_orgao*), ano do exercício da despesa (*ano_exercicio*), categoria de despesa (*ds_elemento*), soma do valor das despesas do agrupamento realizado ($SUM(vl_despesa)$).



Ilustração 1: Elaboração própria

É evidente que a limitação computacional que motivou primariamente essa decisão de agrupamento não é absoluta, havendo a possibilidade de utilização de processamento em nuvem que supere essa restrição. No entanto, não havia disponibilidade desses recursos ao tempo de elaboração desta Dissertação.

De toda forma, a agregação realizada traz informações bastante robustas para a detecção de anomalias e úteis ao procedimento de auditoria. Em verdade, uma avaliação individual de despesas capturaria ruído em excesso, sendo a interpretação de seus resultados de difícil execução e tenderia a gerar uma menor precisão da predição do modelo construído. Na medida em que os dados são agregados, há um natural aumento da precisão do modelo, de modo que desvios percentuais entre o realizado e o predito passam a ser indicativos mais de uma anomalia relevante, ao mesmo tempo em que são mais representativos em termos de valor absoluto de prejuízo.

Por outro lado, é evidente que o tratamento agrupado traz limitações à análise. Em primeiro lugar, características específicas de despesas individuais não podem ser agrupadas, tais como as relativas à contraparte (*identificador_despesa*, *ds_despesa*). A perda dessas referências é particularmente relevante na medida em que a análise de contraparte e de conexões diretas e indiretas com os agentes de decisão do órgão responsável pela despesa é um importante elemento de avaliação de riscos e de anomalias de despesas.

Outro elemento relevante é que a agregação anual traz limitações para a utilização de controles mensais que seriam capazes de recuperar características de sazonalidade das despesas que auxiliariam significativamente a parametrização pretendida. Mais precisamente, para fins de sazonalidade, seria útil ter marcações diárias, dado que determinados tipos de despesas tendem a ser concentradas em datas específicas, como, por exemplo, as de folha de pagamento.

Ainda sobre os aspectos temporais, como típico no processamento de grande quantidade de dados, há um claro *trade off* entre a extensão temporal dos dados utilizados e do aprofundamento da granularidade de cada período. Assim, aqui também tivemos de tomar decisões relevantes. Nesse sentido, seria possível não utilizar a integralidade de dados disponibilizados, restringindo a análise aos períodos mais recentes, e permitindo a utilização de maior quantidade de atributos. No entanto, ao fazê-lo, perderíamos acabar com um viés derivado de uma tendência de gastos ou de perfil de despesas de um determinado período, particularmente quando consideramos os ciclos de renovação eleitoral.

Ademais, conforme desenvolveremos adiante, os atributos que não puderam ser considerados por restrição computacional constituem variáveis categóricas com dezenas, por vezes centenas de categorias. Assim, a sua utilização e transformação em *dummies* requereria uma limitação excessiva do número de exercícios analisados, o que aprofundaria o ponto acima levantado de diminuição de variabilidade de tendências temporais.

Dessa forma, decidimos pela utilização da integralidade dos dados das despesas disponíveis dos 644 municípios paulistas entre os exercícios de 2010 e 2019. Nesse passo, tivemos de escolher, entre os atributos das despesas disponíveis na base de dados da AUDESP, aquelas que seriam utilizadas em nosso modelo preditivo.

Nesse passo, de partida é possível verificar que determinados atributos são de natureza meramente administrativa, e, portanto, não fornecem informações úteis a predições das despesas dos municípios, tais como: *id_despesa*, *nr_empenho*, *cd_programa*, *cd_acao*. De outra parte, há atributos que não são compatíveis com um agrupamento anual das despesas, como os dados de contraparte (*identificador_despesa*), e de identificação de data específica (*mês_referencia*, *mês_referencia_extenso*, *dt_emissao_despesa*).

Um dos atributos cuja retirada dos dados da modelagem gera uma perda relevante de informação é o de *ds_modalidade_lic*. A sua inserção gerava um aumento de 68% dos dados processados. Tal atributo informa qual procedimento legalmente previsto foi utilizado para a aquisição do produto ou serviço. Na medida em que as modalidades de compra são determinadas legalmente, e muitas vezes limitadas por valores máximos em cada modalidade, tal atributo tenderia a diminuir os erros de predição. Ademais, conforme é intuitivo e bem demonstrado por Gallego e Rivero (2021), os atos ilícitos tendem a ser concentrados no âmbito das modalidades de aquisição que conferem maior grau de discricionariedade ao agente público, como nos casos de dispensa ou inexigibilidade de licitação.

Outro atributo que não foi considerado foi o de fonte de recursos. Este refere-se à determinação da origem do recurso que suporta a despesa, e se divide em 14 categorias, abarcando fontes como: recursos próprios da administração, Tesouro (União), transferências e convênios federais, operações de créditos, dentre outras. Aqui também entendemos ser um atributo de interesse para a modelagem do comportamento de despesas, particularmente quando

consideramos que a depender da fonte de recurso, pode haver maior ou menor grau de monitoramento.

Nesse sentido, é interessante marcar que transferências federais são passíveis de verificação pela Controladoria Geral da União – CGU, e que credores como o Banco Nacional de Desenvolvimento Econômico e Social – BNDES ou credores internacionais usualmente impõem suas próprias fiscalizações. Dessa forma, seria de interesse avaliar a associação da fonte de recursos e o comportamento das despesas. Não obstante, isso implicaria realizar agrupamentos adicionais dos dados que esbarram em nossa capacidade de processamento, razão pela qual não utilizamos tal atributo em nossa modelagem.

Um dos aspectos que se mostrou fundamental para a performance do modelo desenvolvido é o que se refere à escolha no nível de classificação de despesa que seria utilizado. Como fica claro na Tabela 1, as despesas são classificadas em diferentes níveis de generalização e com distintos enfoques. É inviável a utilização de todas essas categorias, a uma, por restrição computacional, a duas, por haver um alto grau de colinearidade entre essas categorias, de modo que a sua manutenção conjunta não estaria fornecendo informação relevante para o modelo preditivo. Nesse sentido, foi necessário escolher qual categoria utilizar.

Os atributos que constituem classificação das despesas são: *ds_função_governo*, *ds_subfuncao_governo*, *ds_acao*, *ds_programa* e *ds_elemento*. Há diferenças substanciais entre eles em termos de valores únicos que cada um representa. Na medida em que cada valor único tem de ser transformado em variável *dummy*, um excesso de valores únicos inviabiliza a adoção de determinado atributo na modelagem. Por outro lado, um número excessivamente reduzido de valores únicos deixa de capturar variações fundamentais no comportamento das despesas, reduzindo o poder explicativo do modelo.

Uma técnica usualmente apresentada para abordar o problema apresentado é a análise de componentes principais (*principal component analysis* – PCA). Tal método realiza uma transformação linear ortogonal dos dados, de modo a modificar o sistema de coordenadas para capturar a maior variância na variável de interesse. Dessa forma, a utilização de PCA permite uma simplificação dos dados, reduzindo o número de atributos considerados em favor de um aumento da capacidade de compreensão e da capacidade de processamento.

Uma técnica semelhante para lidar com a questão da dimensionalidade é a *multiple correspondance analysiss* (MCA). Trata-se de uma extensão da análise de correspondência que permite analisar o padrão de relações de variáveis categóricas. Assim, a MCA pode ser visto como uma generalização da análise de componentes principais para variáveis categóricas, sendo mais apropriada para o nosso problema de seleção dos principais atributos de classificação.

Não obstante conhecermos tais técnicas de redução de dimensionalidade, optamos por uma abordagem mais qualitativa na seleção dos atributos de classificação de despesas mais relevantes. É que, para além de uma otimização de critérios de informação, buscamos a criação de um modelo cujo resultado seja útil ao processo de auditoria. Nesse sentido, teve-se por parâmetro o agrupamento das despesas por critérios relacionados a um maior grau de objetividade no que se refere à própria despesa, e menos a vinculações desta a programas ou ações governamentais. Ao fazê-lo, também consideramos a quantidade de valores únicos de cada atributo.

Nesse sentido, entendemos que o melhor atributo para a classificação das despesas seria o *ds_elemento*. Em primeiro lugar, é possível realizar um filtro inicial em razão da quantidade de valores únicos de cada atributo, conforme tabela abaixo:

Tabela 2: Valores únicos das classificações das despesas (exercício de 2015)

Atributo de classificação	Valores Únicos
ds_funcao_governo	28
ds_subfuncao_governo	102
ds_acao	37033
ds_programa	11434
ds_elemento	558

Elaboração própria.

De partida, portanto, é possível depreender a inviabilidade de utilização das categorias *ds_acao* e *ds_programa*, ambas superando dez mil valores únicos apenas no exercício de 2015. Essa imensa quantidade de valores não só resulta em uma incapacidade de processamento, mas é também resultante de uma natureza volátil dos critérios, sendo ambos os casos classificações vinculadas a programas ou ações governamentais particulares de cada

administração e com baixo grau de permanência, não sendo útil a criação de modelos que sirvam a previsões fora da amostra.

Nesse sentido, o ponto de escolha qualitativa se deu, em verdade, entre os atributos *ds_função_governo*, *ds_subfuncao_governo* e *ds_elemento*. Os dois primeiros critérios geram classificações mais abstratas, tendo como vantagem o menor custo computacional, uma diminuição de granularidade dos dados, com uma provável melhor performance do modelo preditivo.

São exemplos de classificações realizadas pelo atributo *ds_funcao_governo*: saúde, cultura, urbanismo, segurança. Já no que se refere a *subfuncao_governo*, podemos citar: ensino profissional, refinanciamento de dívida, serviços urbanos, policiamento. Tais critérios são de grande relevância, e, a depender do uso previsto, poderiam ser os mais apropriados. Nesse passo, se a esta Dissertação tivesse por escopo a avaliação de impactos sobre políticas e serviços públicos, estes tenderiam a ser os níveis de classificação adequados.

Não obstante, quando temos por objetivo gerar modelos úteis à otimização dos procedimentos de auditoria, entendemos que a adoção do atributo *ds_elemento* é mais apropriada. Isso porque este se refere a características objetivas do tipo de produto ou serviço que está sendo adquirido. A título de exemplo, são classificações postas por este atributo: serviços gráficos e editoriais, combustíveis e lubrificantes automotivos, complementação de previdência, auxílios.

Dessa forma, o agrupamento das despesas em torno deste atributo permite uma comparação entre os municípios em critérios mais objetivos. Ao mesmo tempo, na medida em que gera classificações mais pulverizadas, permite que o auditor analise as categorias de despesas com mais eficiência. Nesse sentido, agregações em classificações postas pela *ds_função*, por exemplo, gerariam apontamentos em categorias excessivamente gerais, como “educação”, o que requereria que o auditor analisasse todas as despesas nesta ampla categoria, com alto custo de auditoria.

Por fim, é relevante marcar que filtramos todas as despesas pelo atributo *tp_despesa*, selecionado apenas aquelas que foram registradas como liquidadas. Do contrário,

estariamos multiplicando as mesmas entradas na medida em que a mesma despesa é potencialmente registrada em múltiplos estágios. Assim, a escolha pelo estágio de despesa liquidada se deve ao fato desta etapa constituir a obrigação de pagamento pelo ente.

Ante o exposto, concluímos a seleção dos atributos disponíveis na base de dados de despesas com aquelas mais relevantes para a modelagem, quais sejam: *ano_exercicio* (ano), *ds_municipio* (município), *ds_orgao* (órgão), *ds_elemento* (categoria_despesa), *vl_despesa*. Por fim, montamos base de dados, com o uso de *sqlite*. Por fim, conectamos a base de dados em *python*, com a biblioteca *sqlite3*.

2.2.1 Tratamento dos atributos selecionados.

Realizamos transformações substanciais em dois atributos dos dados originais, quais sejam: *ds_orgao* e *ds_elemento*. Tais modificações foram realizados tendo por objetivo impedir que diferenças de forma fossem interpretadas computacionalmente com distinções de características, o que geraria excesso de ruído, diminuindo a capacidade dos atributos de parametrizar as despesas.

Nesse sentido, é relevante marcar que o atributo *ds_orgao* gera valores distintos para cada nome de órgão existente em cada um dos municípios. Cabe ressaltar, em primeiro lugar, que a nomenclatura utilizada de órgão é não é rigorosa, na medida em que engloba entes da administração direta e indireta (tal como empresas públicas e fundações)¹³. De outra parte, não existe qualquer tipo de padronização na nomeação destes órgãos e não há classificação em abstrata marcada na base de dados originais desses entes por campo de atuação. Se fossem considerados exclusivamente as distinções dos nomes dos entes, no ano de 2015, por exemplo, teríamos 1752 *dummies* de entes. Ocorre que, evidentemente, para fins de parametrização, o que é relevante é o tipo e campo de atuação do órgão, e não o seu nome.

Para tratar tal ponto, utilizamos a biblioteca de processamento de linguagem natural - *natural language tool kit* (nltk). Com ela, excluímos dos nomes dos entes as palavras de baixo

¹³ Em Direito administrativo, a nomenclatura de órgão é utilizada exclusivamente para segmentações da administração direta.

conteúdo semântico, usualmente referidas como *stop words*, como preposições, e, em seguida, listamos as palavras, bigramas e trigramas. De tal forma, conseguimos ter a perspectiva das categorias mais frequentes de atuação dos entes a partir das combinações de palavras de seus nomes. Assim, criamos funções que geraram a classificação de cada entidade em 12 categorias, quais sejam:

- a) Câmara
- b) Prefeitura
- c) Seguridade
- d) Saneamento
- e) Saúde
- f) Cultura/Espportes
- g) Educação
- h) Obra/Construção
- i) Habitação
- j) Outros

Embora tenhamos ciência que essa classificação esteja imbuída de ruídos, ela cria um mecanismo de se extrair informação de áreas de atuação dos municípios e auxilia significativamente na construção do modelo de parametrização das despesas.

Outra transformação dos dados que se mostrou fundamental foi no âmbito do atributo *ds_elemento*. Tal atributo gera como classificação uma *string* em duas partes: código numéricos e nome da categoria de despesa. O código numérico gera uma identificação da categoria da despesa de acordo com o sistema de contabilização do AUDESP. Ocorre que verificamos que a mesma descrição de despesas poderia apresentar codificações distintas e, assim, ser interpretadas pelo modelo como classificações diferentes.

De fato, da perspectiva do sistema de registro do AUDESP, é natural que o mesmo elemento descritivo possa apresentar diferentes códigos de identificação, na medida em que este é construído de modo a vincular a classificação das despesas a outros atributos. No entanto, da perspectiva de elaboração do modelo, os dados seriam interpretados como classificações integralmente distintas apesar de estar tratando de fenômenos em muito semelhantes. Dessa

forma, quando o modelo desenvolvido era treinado para prever os dados agregados sobre esta categoria, o resultado preditivo era muito insatisfatório.

No entanto, tal dificuldade foi resolvida quando aplicamos métodos de transformação dos dados do atributo *ds_elemento*. Segregamos os dados de código e de descrição das categorias de despesa e passamos a utilizar apenas estes. A partir daí, passou a haver entradas com a mesma descrição de atributos. Assim, reagrupamos os dados. Dessa forma, passamos a eliminar o ruído classificatório, o que aumentou expressivamente o resultado preditivo do modelo.

Nesse ponto, é relevante marcar que a equipe técnica do TCE/SP, ao avaliar, preliminarmente, o modelo construído, considerou que uma abordagem alternativa seria possível qual seja, agregar os dados em torno da codificação, retirando a classificação. A relevante ponderação decorre do fato de que a descrição da despesa é sujeita à alteração ao longo do tempo, o que não ocorreria com a codificação.

De fato, em uma eventual implementação operacional do modelo, tal circunstância deve ser levada em consideração. Ocorre que ambas as estruturas de agrupamento foram testadas para os dados disponíveis e não houve alteração relevante da capacidade preditiva. Por outro lado, entendemos que a agregação pela nomenclatura da categoria eleva significativamente a interpretabilidade dos resultados para o público não familiarizados com o sistema de numeração do AUDESP, de modo que preferimos manter esta agregação para fins de apresentação desta Dissertação.

2.3 Características dos municípios.

Para construir uma parametrização das despesas, é necessário incorporar variáveis de controle a nível municipal. Para tanto, utilizamos os seguintes atributos:

- a) População
- b) Produto Interno Bruto (PIB)
- c) Receita líquida (RCL)
- d) Principal atividade econômica

- e) Número de servidores públicos (Vagas)
- f) Número de vereadores (Vereadores)
- g) Área do município (Área)
- h) Número de alunos na rede de ensino (Alunos)

Os dados de população, de PIB, e de principal atividade econômica foram obtidos em base de do IBGE. Já os dados dos demais atributos foram disponibilizados diretamente pelo TCE-SP a partir do sistema interno de dados.

Em primeiro lugar, cabe registrar uma restrição relevante de dados que potencialmente impacta a performance do modelo e a decisão do escopo temporal considerado. Não encontramos base de dados que capturasse variações anuais desses atributos dos municípios do Estado de São Paulo. As avaliações do IBGE ou do Ipea, a título de exemplo, são realizadas muito mais recorrentemente nas capitais dos Estados. Entre os dados disponibilizados pelo TCE, apenas obtivemos dados de alunos entre os anos de 2016 e 2019, e de número de servidores públicos no exercício de 2019.

Os dados de vereadores, área, e principal atividade econômica variam entre os municípios, sendo um relevante atributo de parametrização, mas, para um mesmo município, são virtualmente constantes ao longo do tempo. Dessa forma, para estes, entendemos que não há qualquer prejuízo em manter os atributos fixos por município.

Por outro lado, seria relevante ter obtido dados anuais de população, número de servidores e de alunos. Tais dados podem sofrer relevante variação ao longo do tempo e possuem o potencial de gerar significativa capacidade explicativa da variância do valor das despesas. Na medida em que, conforme ressaltado, para o período mais recente, tínhamos acesso à variação dos dados de número de servidores e de alunos, poderíamos ter optado por restringir o alcance temporal do modelo. Isso, contudo, implicaria em perda de informação relevante das despesas, podendo enviesar o resultado para capturar uma tendência próximo ou determinado ciclo político.

Dessa forma, optamos por utilizar os dados de PIB, População, alunos matriculados, e número de servidores como fixos para cada município, tendo por base os dados de 2019.

Embora haja uma perda do poder explicativo, tais dados ainda representam uma *proxy* dos dados que seriam obtidos em uma avaliação anual.

Dentre estes atributos que fixamos, a característica que está mais suscetível a mudanças significativas é o PIB. No entanto, conseguimos dados anuais da receita líquida dos municípios que captura em grande medida os choques econômicos. A decisão de manter o PIB e a receita líquida simultaneamente como várias explicativas apenas se justifica na medida em que a receita do município não é afetada, pela mesma intensidade por todos os setores econômicos, sendo relevante a manutenção de informação sobre uma *proxy* de relação do PIB e receita líquida.

Nesse passo, é relevante lembrar que a arrecadação dos municípios é concentrada pelas receitas de ISS e de IPTU. Dessa forma, municípios com um mesmo PIB pode ter diferentes níveis de receita. Por outro lado, a relação de descompasso entre PIB municipal e receita líquida pode ser um fator relevante para os desafios da gestão municipal, razão pela qual optamos por manter ambas as variáveis como controle das despesas municipais.

A variável de população é fundamental na parametrização das despesas, indicando, per si, a pressão por despesas municipais. Quando associada à área, traz elementos de densidade populacional que também é fundamental na determinação de gastos municipais como, por exemplo, de combustível, transporte, dentre outros. E, quando controlado pela RCL, traz o indicativo de recursos per capita, que também é um elemento essencial na explicação do nível de gastos.

A variável de número de servidores serve como um parâmetro óbvio do gasto com servidores públicos, mas também é um elemento de referência de gastos administrativos em geral. Pode ser argumentado também que o nível de gastos com os servidores também pode estar associado à capacidade da burocracia (em sentido weberiano), e, também, auxiliar a parametrização de despesas por esse canal. De forma semelhante, o número de vereadores é um claro elemento de controle do nível de gastos associados às atividades das câmaras de vereadores e de ações parlamentares.

Ainda, o número de alunos é utilizado para parametrizar essa relevante fração dos gastos municipais que é a educação. Os dados fornecidos pelo TCE-SP são subdivididos em categorias como creche, pré-escola, anos iniciais e anos finais. No entanto, para fins de construção de modelo, agregamos todas essas faixas.

Seria fortemente argumentável a decisão de apenas utilizar os anos de educação que correspondem ao ensino fundamental, que é o campo de obrigação constitucional municipal. Ocorre que muitos municípios oferecem cursos técnicos, de formação profissional ou mesmo faculdades. Ainda, o número total de alunos carrega informações adicionais, indicando um nível mais geral de escolaridade que pode ser determinante na diferenciação do comportamento dos municípios.

Ainda, a variável categórica indicativa de principal atividade econômica na composição do PIB municipal. O IBGE prevê as seguintes categorias:

- a) Indústrias extrativistas;
- b) Comércio e reparação de veículos automotores e motocicletas;
- c) Pecuária, inclusive apoio à pecuária;
- d) Eletricidade e gás, água, esgoto, atividades de gestão de resíduos e descontaminação;
- e) Indústrias de transformação;
- f) Agricultura, inclusive apoio à agricultura e pós-colheita;
- g) Demais serviços.

Ao utilizar essa categorização em nosso modelo, conseguimos ter um indicativo das diferentes organizações socioeconômicas que podem explicar comportamentos distintos das despesas dos municípios.

2.4 Análise Descritiva.

Neste tópico, passaremos a realizar a avaliação descritiva dos dados que serão utilizados para modelagem. Faremos a exposição gráfica dos atributos de modo a construir uma percepção geral dos dados, o que auxilia a análise de decisões relevantes de tratamento de dados e de interpretação dos resultados.

De partida, apresentamos um gráfico de associação entre os atributos utilizados. A elaboração deste gráfico requer cuidado especial na medida em que, entre as *features* utilizadas, há variáveis de natureza numérica, mas também outras de natureza categórica. Para avaliar, portanto a associação entre variáveis categóricas, utilizamos a biblioteca de *python* para o processamento de dados categóricos *dython*. Embora fuja ao escopo da presente dissertação aprofundar os fundamentos utilizados pela biblioteca, é relevante pontuar que o algoritmo é baseado na medida de correlação de variáveis discretas Cramer's V.¹⁴

Dessa forma, o gráfico abaixo apresentado indica o grau de associação entre as variáveis do modelo. Apesar de não aplicarmos efeitos fixos de municípios em nosso modelo, optamos por inseri-los no gráfico de correlação para fins apresentação da sua relação com as variáveis de interesse.

¹⁴ LIEBETRAU, Albert M. *Measures of association*. Newbury Park, CA: Sage Publications. *Quantitative Applications in the Social Sciences Series* No. 32., 15–16p.

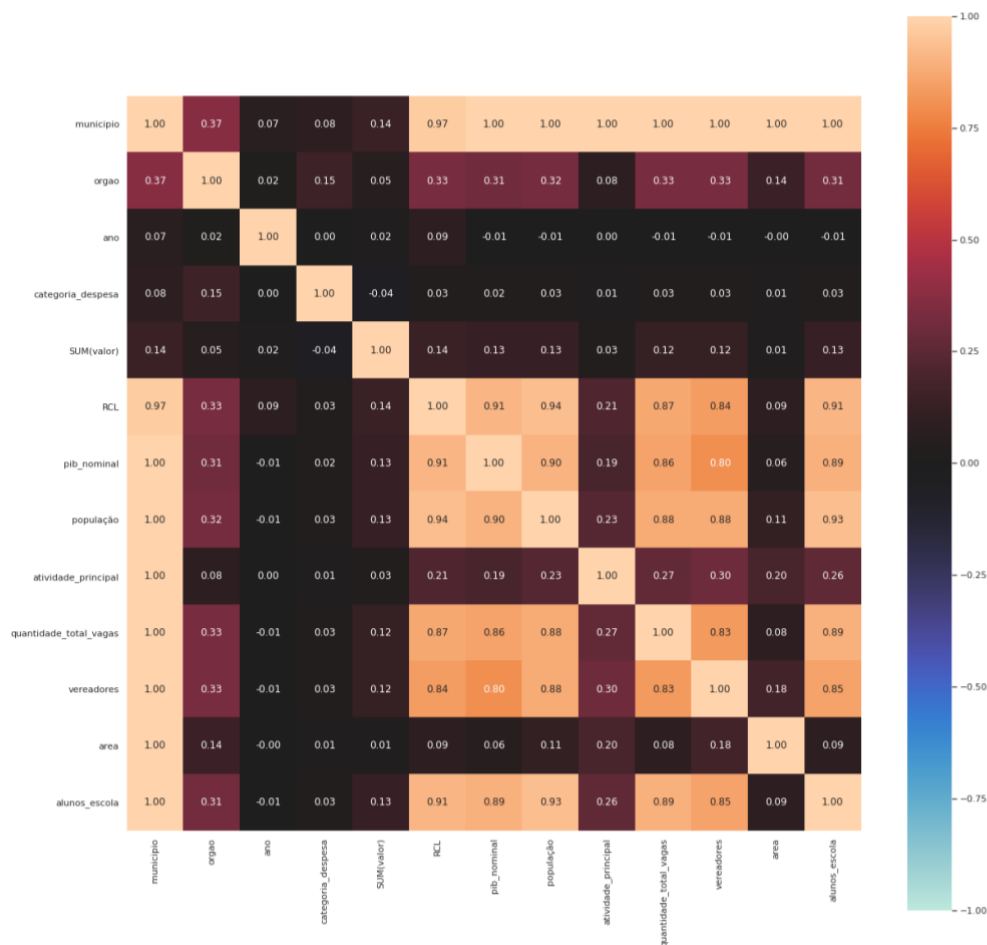


Gráfico 1: Elaboração própria.

Neste gráfico, verificamos o alto grau de correlação entre os municípios e as suas características, o que é, por óbvio, esperado, particularmente quando, com a exceção de RCL, os dados são fixados no tempo. Apontamos esse fato, contudo, para a apresentar uma preocupação de que, apesar de não estar utilizando efeitos fixos de município, o modelo naturalmente acabe capturando padrões de comportamento de cada município.

A decisão de não utilizar efeitos fixos de municípios é motivada pelo objetivo de prever despesas esperadas por fatores explicativos objetivos e não por razões idiossincráticas. Nesse sentido, se determinado município é recorrentemente ineficiente ou corrupto, não queremos que essa característica seja considerada de modo a corrigir o valor esperado considerando esse comportamento. Ao contrário, gostaríamos que houvesse um erro significativo entre o resultado esperado pelo modelo e o efetivamente realizado pelo município ineficiente/corrupto.

Dessa forma, tínhamos a preocupação de que os métodos de *machine learning*, particularmente os que utilizam métodos de segmentação dos dados como *decision trees*, acabassem por gerar estimativas que capturassem indiretamente as idiossincrasias de cada município. Tal ponto será abordado na quarta seção desta Dissertação.

Outro aspecto de destaque é a associação do valor das despesas e as *features* do modelo. Verifica-se que as características dos municípios possuem o maior grau de associação, mas nenhuma com um peso muito significativo individualmente, variando com uma associação entre 0.11 e 0.14. Não é capturado, pelo uso do *dython*, associação significativa entre a categoria de despesas e o valor das mesmas. No entanto, é preciso considerar que a associação calculada tem por base uma avaliação de significância estatística de variável que representa 558 valores únicos, o que requer um poder de teste excessivo. No mais, *dython* busca relações lineares, e, conforme desenvolveremos adiante, o nosso modelo é, essencialmente, não linear.

a) Valor das Despesas.

Na medida em que esta será a variável dependente dos modelos construídos, construíamos a tabela abaixo com algumas estatísticas relevantes das variações do valor das despesas que servirão de referência relevante:

Tabela 3: Estatísticas – Valor de despesa

Máximo	1,395 bilhões
3° quartil	152 mil
mediana	19 mil
1° quartil	3 mil
Mínimo	2
média	830 mil
Desvio padrão	9,288 milhões

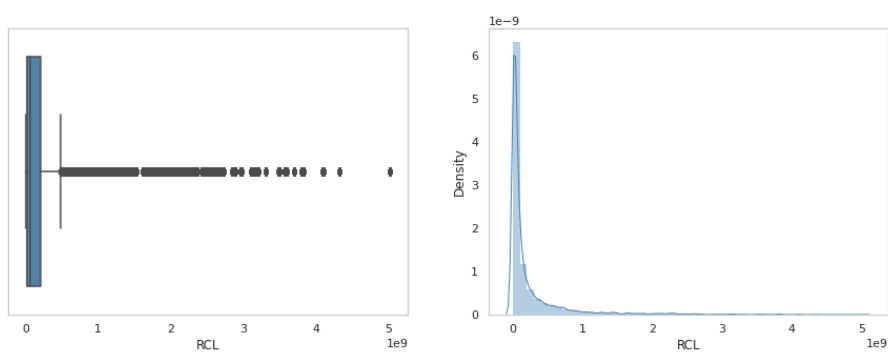
Fonte: Elaboração própria.

Há variações entre as categorias de despesas em um mesmo município e entre municípios. Considerando esse aspecto, mostraremos, no próximo capítulo, que optamos, em

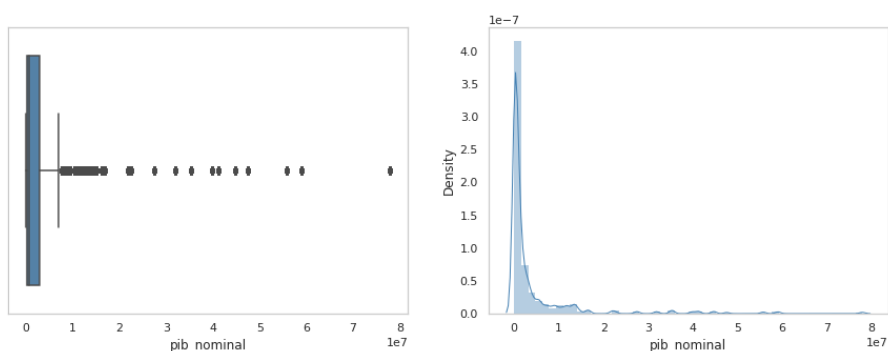
um dos modelos, por realizar clusterizações por município, e, em outro, por nível de despesas, treinando um modelo por cluster.

Ademais, foi verificada a existência de entradas negativas da base de dados, que representariam estornos de valores lançados equivocadamente. Ocorre que, muitas vezes, esses estornos são realizados em exercícios distintos da despesa original, o que gerava, inclusive, dados de valores agregados negativos em determinadas categorias. Dessa forma, optamos por filtrar todos os valores negativos da agregação em *sqlite*. Isso gera valores sobrestimados. No entanto, é razoável assumir que os erros que requerem estorno são aleatórios.

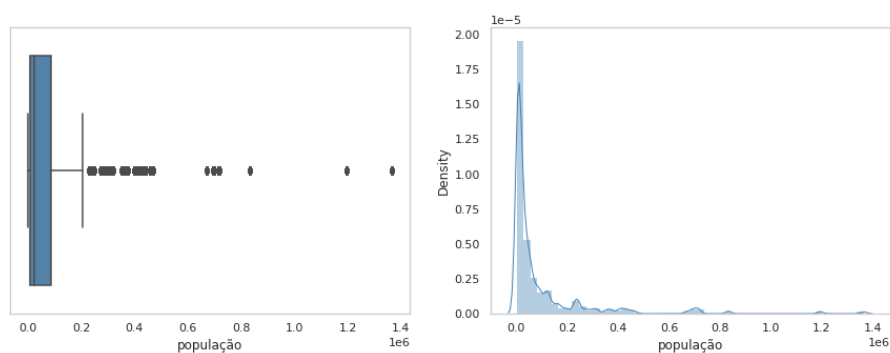
b) RCL, PIB e população.



Gráficos 2: Elaboração própria.



Gráficos 3: Elaboração própria.



Gráficos 4: Elaboração própria.

As distribuições apresentadas nos gráficos acima de RCL, PIB e população deixam patente que os municípios apresentam características altamente desbalanceadas em tais atributos. Ao mesmo tempo, conforme já adiantado no Gráfico 1, tais distribuições possuem alta correlação entre si. Tais fatos sustentam a decisão, apresentada a frente, de testar modelos específicos por *clusters* de municípios.

c) Contagem de categorias.

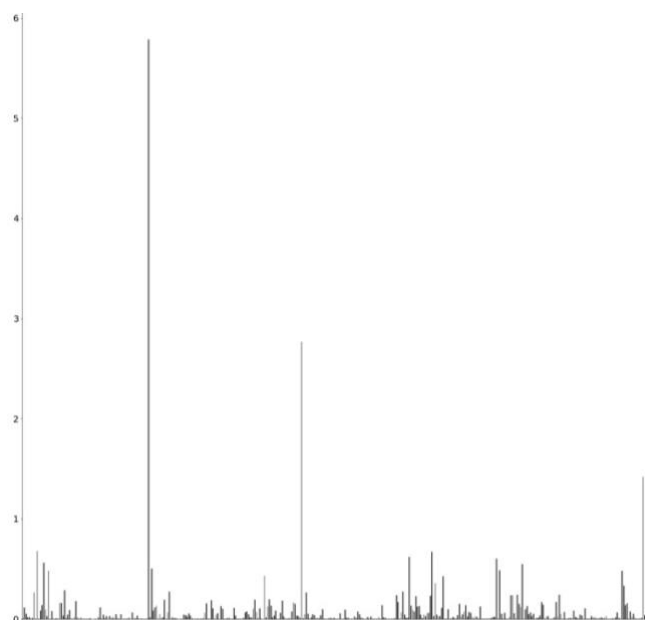


Gráfico 5: Elaboração própria.

Apenas com o fim de dimensionar a distribuição das categorias de despesa em nosso *dataset*, apresentamos o gráfico acima mostrando quão frequentes são as categorias. É relevante

marcar que estamos tratando de agregações anuais. Ademais, ressaltamos que não estamos apresentando o valor das despesas por categoria, mas a sua simples frequência. Percebe-se, pois, que há categorias com uma frequência relativa muito maior, o que deve ser levado em consideração na modelagem.

d) Contagem de entradas por município.

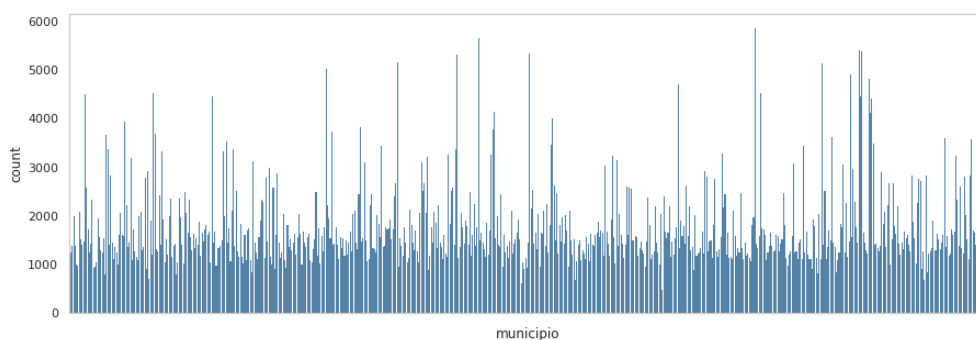


Gráfico 6: Elaboração própria

Por fim, registre-se que, conforme mostra o gráfico acima, embora a maior parte dos municípios apresente o mesmo número de entradas, há aqueles que se destacam por um significativo maior número. Quando lembramos que todas as despesas são agrupadas anualmente, por órgão e por categoria de despesa, essa discrepância passa a ser explicada pelo fato de determinados municípios apresentarem uma maior variedade de tipos de entidades municipais em sua administração ou por apresentarem um maior número de categorias de despesas usualmente realizadas. Tal fato deve ser levado em consideração, pois pode gerar um maior peso do perfil de despesas desses municípios (ou por choques ocorridos no âmbito destes) na parametrização de despesas gerada pelo nosso modelo.

Uma vez tendo sido realizada a análise descritiva dos dados disponíveis para modelagem, e tendo sido dada a devida transparência dos fundamentos das escolhas de variáveis, assim como das decisões de agregação dos dados, passaremos, na seção que segue, a expor os modelos de parametrização construídos e testados.

3 Modelos e Resultados.

Nesta seção, passaremos a descrever os modelos desenvolvidos para a parametrização das despesas municipais. Tais modelos foram construídos com o propósito último de auxiliar os procedimentos de auditoria. Espera-se que a parametrização possa fornecer referências robustas aos auditores de ocorrência de despesas anormais.

Com isso, o processo de controle tende a não só ampliar a eficiência em termos de custos, com a concentração dos esforços dos auditores em contas com a maior probabilidade de irregularidades ou ineficiências materiais, como também a trazer melhoras qualitativas, incrementando a detecção de anormalidades que poderiam não ser percebidas em uma avaliação individual das contas dos municípios.

De uma perspectiva mais geral, queremos diminuir os custos de monitoramento e aumentar a probabilidade de detecção de irregularidades, impactando na decisão do agente de cometimento de ilícitos, e aumentando o seu ponto ótimo de esforços em torno de uma boa gestão. Ao fim, um bom modelo de parametrização de despesas pode servir a uma diminuição da assimetria de informações entre sociedade e agente público.

A essência de todos os modelos que serão apresentados é a utilização do comportamento histórico das despesas, controlados por variações de características relevantes dos municípios, e por choques temporais (efeitos fixos de ano) para estimar o que seria esperado de determinada conta de órgão municipal.

De partida, apresentaremos os resultados de um modelo linear simples. Os objetivos deste modelo são dois: (i) servir de *baseline*, uma referência do desempenho dos demais modelos; (ii) demonstrar que não estamos diante de um problema linear, o que justifica a adoção de modelos de *machine learning* que sejam capazes de capturar tais relações não lineares.

Em seguida, passamos a apresentar os resultados da utilização do *Gradient Booster Regressor*. Tal algoritmo insere-se dentro da categoria de *Ensemble*, que são aqueles que utilizam de conjunto de regressores para a estimação. Tal algoritmo de *machine learning* adiciona sequencialmente estimadores, cada um corrigindo os erros residuais de seus

predecessores. Assim, este é usualmente um excelente mecanismo para gerar regressões não lineares.

Depois, apresentamos os resultados do modelo de *Random Forest Regressor*. Tal algoritmo é um *ensemble* de árvores de decisão (*Decision Trees*).

Seguimos, no item subsequente, apresentando os resultados de uma rede neural construída com o mesmo objetivo. Não pretendemos ser exaustivos quanto a este tópico, que foi apenas elaborado com o objetivo de avaliar as possibilidades de utilização desse tipo de arquitetura neste tipo de problema. Como veremos, embora tenhamos obtido resultados promissores, estes não superam a performance do *Random Forest Regressor*, nem do *Gradient Booster Regressor*.

Por fim, uma vez tendo passado pelos modelos testados, e demonstrando a melhor performance, passamos, em item em apartado, a explorar mecanismos de aprimoramento que possam gerar uma diminuição de erros verificados no modelo construído. Nesse sentido, são apresentadas duas abordagens: (i) clusterização dos municípios pelas suas características, gerando um modelo preditivo por *cluster*; (ii) segregação das despesas por quartil da distribuição de valores de despesas, gerando um modelo preditivo por quartil.

Os resultados que serão apresentados foram todos produzidos usando bibliotecas em Python. Mais especificamente, o modelo linear, *Random Forest Regressor*, e o *Gradient Booster Regressor* foram todos derivados da biblioteca *scikit-learn*. Já os modelos em redes neurais foram construídos utilizando *Tensor Flow/Keras*.

Foge ao escopo desta Dissertação a descrição dos fundamentos teóricos de todos os modelos de *machine learning* apresentados, tendo sido adotada uma abordagem estritamente aplicada destes com o propósito de contribuir para a literatura empírica de detecção de corrupção com a criação de um instrumento de parametrização de despesas públicas municipais.

Todos os modelos apresentados terão por objeto estimar o valor das despesas de determinado órgão municipal em cada categoria de despesas agregadas por ano (*vl_despesa*), e para tanto, utilizarão as mesmas variáveis explicativas (*features*), quais sejam:

- a) Categoria de despesa (transformada em variáveis *dummies* | *one-hot-ecoding*)
- b) Órgão municipal (transformada em variáveis *dummies* | *one-hot-ecoding*)
- c) Ano (transformada em variáveis *dummies* | *one-hot-ecoding*)
- d) População municipal
- e) PIB municipal
- f) Receita líquida municipal
- g) Principal atividade econômica do município (transformada em variáveis *dummies* | *one-hot-ecoding*)
- h) Número de servidores públicos do município
- i) Número de vereadores do município
- j) Área do Município

Como forma de garantir generalidade dos modelos criados, evitando problemas de *overfitting*, em primeiro lugar, separamos, por seleção aleatória, 20% dos dados, para verificação final dos resultados do modelo selecionado.

A avaliação de cada modelo, bem como a comparação entre eles, é gerada pela separação aleatória de grupos de treinamento e de validação com o uso de *cross validation* em 5 *folds*. Assim, os resultados da próxima seção são o resultado médio dos cinco *folds* na amostra de validação. Ao final, apresentaremos os resultados no âmbito da validação na amostra de teste.

Ainda, é relevante marcar que sempre geramos padronizações dos dados, como é necessário em particular no uso de redes neurais, utilizamos apenas a escala da amostra de treino de modo a evitar uma contaminação indireta do processo de treinamento com informação da amostra de teste. Tais padronizações foram realizadas pelo uso de *StandardScaler*. Por fim, registre-se que os hiperparâmetros dos modelos (a exceção das redes neurais) foram ajustados com o uso de *GridSearchCV*.

Todos os resultados serão apresentados em três métricas: R^2 , erro médio absoluto e erro mediano absoluto. Sem uma modelagem de parametrização, se o processo de auditoria tivesse de buscar parâmetros de comparação de gastos entre os municípios, um procedimento

naive seria simplesmente comparar em relação à média dos gastos. Para ter esse cenário como referência, apresentamos aqui quais seriam os resultados dessa estratégia com essas três métricas:

Tabela 4: Modelo Naive I - Erros em relação à média.

R^2	Erro médio absoluto	Erro mediano absoluto
0	1.347.087	817.296

Fonte: Elaboração própria

Outra possibilidade de referência *naive* seria sempre utilizar a mediana dos gastos como parâmetro. Nesse caso, teríamos os seguintes resultados:

Tabela 5: Modelo Naive I - Erros em relação à mediana.

R^2	Erro médio absoluto	Erro mediano absoluto
-0,7%	824.481	18.870

Fonte: Elaboração própria

3.1 Modelo Linear.

Com um propósito primário de criar referência mais robusta de performance do processo de modelagem quando comparado com a referência *naive* acima apresentado, geramos um modelo linear simples. Considerando que o objetivo deste era servir de *baseline*, não foram realizados quaisquer ajustes de hiperparâmetros.

Tabela 6: Modelo Linear – Resultados

R^2	Erro médio absoluto	Erro mediano absoluto
7,87%	1.651.084	751.159

Fonte: Elaboração própria

Verifica-se um baixíssimo por explicativo do modelo linear, apresentando um R^2 de 7, 87%. Tal resultado demonstra, de plano, a dificuldade do problema que se coloca. Ao mesmo tempo, demonstra-se que a utilização de modelos lineares é de baixa utilidade ao nosso desiderato, o que motivou a construção de modelos de *machine learning* que são mais eficientes não captura de não linearidades.

3.2 Gradient Booster Regressor.

Para a construção do modelo de *Gradient Booster Regressor*, utilizamos como controle de hiperparâmetro apenas a limitação de profundidade máxima (*max_depth*) em 100. Tal controle foi resultado de usos de mecanismos de *GridSearch*. Limitações computacionais impediram que todos os parâmetros fossem testados em um *range* excessivamente aberto. Dessa forma, a limitação da profundidade serviu, principalmente, a evitar um *overfitting*, que geraria a perda de generalidade e, portanto, de desempenho do modelo treinado fora da amostra de treino.

A maior expectativa de performance, de partida, girava em torno do algoritmo *Gradient Booster Regressor* na medida em que este usualmente supera os resultados do *Random Forest* (Geron, 2019). De fato, obtivemos um alto nível de performance, particularmente quando considerados os resultados do modelo *baseline*:

Tabela 7 : Gradiente Booster Regressor - Resultados

R^2	Erro médio absoluto	Erro mediano absoluto
88,93%	320.898	15.912

Fonte: Elaboração própria

Verificamos, pois, que obtivemos um R^2 de 89,93%, um resultado mais de onze vezes superior ao modelo *baseline*. Assim, conseguimos, com o uso deste modelo treinado, explicar quase 90% dos desvios em relação à média. Quando lembramos da imensa variância das despesas, essa capacidade de explicação dos desvios é um excelente resultado.

Considerando a finalidade de criação do modelo, passa a haver um parâmetro de comparação dos gastos entre os municípios que serve de instrumento de avaliação das despesas individuais muito superior a um mero uso de média dos gastos, da sua mediana, ou mesmo de um modelo linear.

De outra parte, verifica-se a existência de um erro médio absoluto de 320 mil (cinco vezes menor do que o modelo linear). Em números absolutos, trata-se de um erro alto. Mesmo quando comparado com a média dos gastos, tal erro representa 38% da média, o que é muito

significativo. No entanto, é preciso considerar que o desvio padrão é de 9,288 milhões. Assim, o erro médio absoluto representa apenas 3,44% do desvio padrão.

Ainda, o erro mediano absoluto é de 15.912, sendo que a mediana dos gastos é próxima a 19.000. Tal erro mediano é ainda bastante significativo, sendo 83% do valor da mediana das despesas. No entanto, representa uma melhora de 15% relação a uma relação a uma simples referência a mediana (Modelo Naive II).

3.3 *Random Forest Regressor*.

Para a construção do modelo usando *Random Forest Regressor* também utilizamos GridSearch para a otimização de hiperparâmetros e limitamos a profundidade máxima em 100, mantendo os demais em *default*. Obtemos os seguintes resultados médios em *cross validation* em 5 folds:

Tabela 8: Random Forest Regressor - Resultados

R^2	Erro médio absoluto	Erro mediano absoluto
91,60%	322.923	26.551

Fonte: Elaboração própria

Percebemos que o modelo em *Random Forest*, quando comparado ao uso de *Gradient Booster Regressor*, apresenta um aumento de performance em termos de R^2 , sendo possível explicar 91,60% dos desvios em relação à média, o que representa um incremento de 3% da capacidade explicativa.

Por outro lado, a melhora no desempenho em R^2 vem ao custo de uma piora tanto em termos de erro absoluto médio quanto de erro absoluto mediano. Em termo de percentuais, há um aumento de 0,6% do primeiro, e de 66% do erro mediano absoluto.

Não obstante haja uma melhora em termos de R^2 , a expressiva piora no erro mediano é excessivamente grande. Em particular, há de se notar que erro mediano é maior do que seria alcançado pelo Modelo Naive II. Assim, estivéssemos diante de uma simples escolha entre o *Random Forest Regressor* e *Gradient Booster Regressor*, haveria elementos para a

escolha do último. Ocorre que, como veremos adiante, a clusterização por quartil de despesa diminui em grande parte os erros medianos do *Random Forest Regressor*.

3.4 Rede Neural.

Também foi testada a performance de redes neurais para a predição das despesas. Testamos diversas arquiteturas em camadas sequenciais densas com uso de *Tensor Flow* e *Keras*. O modelo de melhor performance utilizou 11 camadas densas com 1024 neurônios cada e ativação *relu*, seguida de uma camada linear de previsão. Foi utilizado o otimizador *RMSprop* com uma taxa de aprendizado de 0.001. Ainda, usamos como função de perda o erro quadrático médio. Seguem os resultados:

Tabela 9: Rede Neural - Resultados.

R^2	Erro médio absoluto	Erro mediano absoluto
86,94%	413.598	25.519

Fonte: Elaboração própria

Percebe-se que há um R^2 um pouco inferior ao encontrado com o uso do *Gradient Booster*, um erro médio absoluto 29% maior e um erro mediano absoluto 60% maior.

3.5 Comparação Geral.

Para facilitar a comparação de todos os modelos expostos, apresentamos a adiante a tabela comparativa geral:

Tabela 10: Resumo final de comparação

	R^2	Erro médio absoluto	Erro mediano absoluto
Modelo Naive I	0	1.347.087	817.296
Modelo Naive II	-0,7%	824.481	18.870
Modelo Linear	7,87%	1.651.084	751.159
Gradient Booster Regressor	88,93%*	320.898 **	15.912 **
Random Forest Regressor	91,60% **	322.923 *	26.551 *
Rede Neural	86,94%	413.598	25.519

Fonte: Elaboração própria

3.6 Considerações intermediárias acerca da performance em *cross validation*.

Na medida em que o resultado da rede neural não supera o *Gradient Booster* em nenhuma métrica, não o consideramos para a modelagem final. Dessa forma, a análise aqui desenvolvida é concentrada na avaliação de performance entre *Random Forest Regressor* e o *Gradient Booster Regressor*.

Ambos os modelos apresentam um R^2 muito alto, acompanhado de um erro mediano absoluto significativo. Tal resultado induz à interpretação que estes modelos geram um alto poder explicativo das grandes variações das despesas, mas tem resultados mais modestos na explicação de pequenas discrepâncias absolutas.

Considerando os objetivos de construção do modelo, os resultados são muito promissores. Com ele, há um mecanismo adequado de parametrização de gastos que serve a informar o processo de auditoria. As grandes variações são explicadas pelas características de despesas e variações absolutas. Dessa forma, se a atividade de controle vier encontrar grandes discrepâncias não previstas pelo modelo construído, haverá um significativo indício de que esta despesa seja uma anomalia. Dessa forma, o modelo serve ao propósito de parametrização das despesas, trazendo um instrumento de referência robusta para o processo de auditoria.

Não obstante, os erros medianos proporcionalmente altos em relação à mediana dos gastos parece uma fragilidade relevante do modelo gerado. Há diferentes explicações possíveis para esse comportamento: (i) elementos idiossincráticos dos municípios; (ii) informações faltantes de variação das características municipais ao longo do tempo; (iii) excessivo desbalanceamento dos dados.

De um lado, propositalmente, não foram utilizados efeitos fixos de municípios. Dessa forma, os modelos construídos utilizam como *features* características municipais, mas não os municípios em si. Essa decisão foi tomada exatamente para que a atividade de controle possa ser informada acerca das diferenças entre o nível de despesas que seria esperado pelo

modelo, considerando fatores explicativos gerais, e as despesas usualmente realizadas por determinado município.

Assim, o modelo preditivo não corrige o valor esperado simplesmente porque determinado município usualmente superfatura determina categoria de despesa, o que aconteceria com a aplicação de efeitos fixos de município. Por outro lado, o rol de elementos de controle utilizados é, evidentemente, limitado. Dessa forma, pode haver diversas características não observadas que sejam motivos adequados de variações entre os gastos municipais. O modelo não consegue capturar tais elementos e, assim, é razoável esperar que variações de baixo valor absoluto não sejam previstas.

Se é verdade que, de partida, já seria de esperar uma menor capacidade preditiva de pequenas variações, quando consideramos as restrições de registros de variações nos dados disponíveis, temos motivos adicionais para tanto. Nesse sentido, lembre-se que, em nossos dados, apenas temos variações anuais de dados de receita líquida dos municípios. Assim, a título de exemplo, se, ao longo dos anos, há alterações do número de alunos matriculados no sistema educacional de municípios, seria esperado que houvesse aumentos das categorias de gastos vinculados à educação naquele município. Tais alterações não seriam muito grandes, mas seriam esperadas. No entanto, como apenas consideramos os dados das características em determinado ano, o modelo não tem capacidade de prever essas pequenas variações.

Por fim, é necessário marcar que os dados de despesas são imensamente desbalanceados, com enorme variância. Isso traz grande dificuldade preditiva de variações em torno de uma média que não é representativa. Nesse sentido, embora os modelos consigam estimar por volta de 90% da variação em relação à média, os desvios absolutos ainda são proporcionalmente muito representativos em relação aos valores realizados de despesas de baixo valor. Diante disso, testamos duas abordagens distintas: (i) clusterização de municípios pelo uso do algoritmo *K Means*; e (ii) segregação dos dados por quartil de despesa. Nos tópicos a seguir, faremos a exposição dessas abordagens.

3.6.1 Clusterização de municípios.

O fundamento da estratégia de clusterização de municípios é, a partir das características municipais, segregar os municípios que se mostrem comparáveis. Nesse caso, um modelo é treinado especificamente para cada cluster de municípios, o que tenderia a diminuir o ruído do processo.

Para a clusterização, agrupamos os dados com nas seguintes características:

- a) Receita Líquida;
- b) PIB;
- c) População;
- d) Número de Servidores;
- e) Número de vereadores
- f) Área
- g) Principal atividade econômica

Em seguida, padronizamos os dados com *StandardScaler*, e utilizamos o *KElbowVisualizer*, da biblioteca *Yellowbrick*. Para determinar o número ótimo de clusters para a divisão dos municípios: treze. Assim, utilizamos o algoritmo Kmeans para agrupar os municípios por cluster.

Uma vez determinada a divisão dos clusters municipais, marcamos cada entrada dos dados de treinamento com seu respectivo cluster, e passamos a treinar um modelo de *Random Forest Regressor* por cluster. A utilização deste em detrimento do uso do *Gradient Booster Regressor* se deu em razão de o modelo em *Random Forest Regressor* ter apresentado um maior R^2 , havendo a possibilidade que o seu uso em dados clusterizados pudesse gerar uma melhora nos critérios de erros médio absoluto e de erro mediano absoluto.

Em termos de hiperparâmetros, para cada modelo, foi testado um range de 5 valores de profundidade máxima entre 20 e 120. Abaixo, apresentamos os resultados médios de um *cross validation* em cinco *folds* de cada um dos modelos, acompanhados, para fins de referência, do valor médio e mediano de cada *cluster*:

Tabela 11: Resultados de modelos desenvolvidos por cluster

Cluster	Número de observações	Valor médio	Valor Mediano	R^2	Erro médio absoluto	Erro mediano absoluto
0	390.827	294.834	15.000	89.18%	111.986	13.450
1	72.979	120.513	9.871	92.57.15%	43.241	7.751
2	90.766	2.589.176	59.995	87.30%	1.103.314	86.669
3	111.142	163.275	11.962	91.96%	57.198	9.630
4	139.734	612.615	19.495	91.29%	236.542	22.756
5	8.246	8.936.183	166.797	93.35%	3.674.683	265.152
6	96.406	533.867	18.558	70.02%	272.482	24.212
7	4.359	898.012	25.974	55.59%	701.583	60.319
8	11.913	269.359	15.784	83.06%	122.479	19.720
9	3.211	199.645	15.132	82.43%	93.651	16.853
10	15.592	716.679	17.308	77.03%	407.719	24.035
11	37.479	4.645778	60.803	87.60%	1.893.443	88.950
12	146.540	1.079.636	33.562	89.17%	434.959	40.356

Fonte: Elaboração própria

O que é possível verificar é que a divisão por clusters de municípios não auxiliou significativamente na redução da proporção do erro mediano e valor mediano. No entanto, é possível verificar alguns padrões relevantes.

Nesse sentido, em primeiro lugar, é possível perceber que, embora a clusterização por municípios diminua as discrepâncias entre valor médio e mediano, ainda há diferenças muito expressivas. Isso induz a interpretação de que os maiores *outliers* ocorrem por diferenças de tipos de despesa, e não apenas por tipos de municípios. Assim, a abordagem que é utilizada na próxima subseção tende a lidar com esse ponto, agrupando as despesas por quartil.

De outra parte, verifica-se que o R^2 por cluster de municípios é distribuído de forma mais ou menos uniforme, apresentando, nessa métrica, valores acima de 80%. Apenas no cluster 7 verifica-se uma queda mais expressiva do R^2 , para 55.59%, sendo ainda bastante razoável.

3.6.2 Clusterização por quartil de valor de despesas.

Diante do exposto, uma outra abordagem realizada visando a melhor capacidade preditiva de valores menores foi segregar os dados de acordo com o quartil das despesas. Para

tanto, após a separação entre amostra de treino e de teste, determinamos as faixas de cada quartil de valor de despesa tomando por base a escala da amostra de treino, marcamos cada entrada como pertencendo a seu respectivo quartil, e passamos a treinar quatro modelos de *Random Forest Regressor*.

Aqui também, para a escolha dos hiperparâmetros, testamos um *range* de cinco valores de profundidade máxima entre 5 e 120. Abaixo apresentamos os resultados médios do *cross validation* em 3 *folds*:

Tabela 12: Resultados dos modelos desenvolvidos por quartil de despesa

Quartil	Valor Médio	Valor Mediano	R^2	Erro médio absoluto	Erro mediano absoluto
1º	1.068	855	10.86%	639	549
2º	8.734	7.680	13.38%	3.328	2.797
3º	63.468	53.289	29.87%	23.622	18.794
4º	3.253.746	566.274	90.87%	1.176.489	226.671

Fonte: Elaboração própria

Aqui, é possível capturar com clareza a essência da dificuldade de modelagem da predição das despesas municipais. Salta aos olhos que o modelo tem uma capacidade de análise muito superior dos grandes desvios em relação à média do que dos menores.

Nesse sentido, é possível perceber que, no âmbito das despesas de menor volume, a distribuição é menos enviesada, com um aumento progressivo da influência de *outliers*, que é acompanhada também por um ganho progressivo de R^2 , até que, no último quartil, a diferença entre valor médio e mediano explode, e, com isso, também o R^2 . Ressalte-se nesse raciocínio que, em que pese a razão entre média e mediana ter sido marginalmente alterada do segundo para o terceiro quartil, o R^2 é dobrado, indicando que, para além da presença de *outliers*, o próprio nível das despesas também é relevante¹⁵.

Se, de um lado, temos a confirmação da restrição na capacidade explicativa dos desvios dos menores valores de despesas, por outro lado, verificamos que a segregação dos

¹⁵ É relevante lembrar aqui que não parece haver indicativo que isso seja resultado de um problema de escala dos dados, pois, previamente à modelagem, todas as colunas foram devidamente padronizadas por *StandardScaler*.

dados por quartil geraram erros medianos significativamente menores. Nesse ponto, o modelo geral de *Random Forest* apresentava uma razão entre erro mediano e valor mediano de 140%. Mesmo o *Gradient Booster* que, nessa métrica, apresentava resultados melhores, tinha essa razão em 83%. Com a agregação por quartil, passamos a ter 63%, 35%, 34.5%, e 38%, em ordem crescente de quartil, o que representa uma melhora expressiva de performance.

Todos os erros médios absolutos foram substancialmente menores que o erro médio. Apesar da performance em termo de R^2 ser menor nos primeiros quartis, é claro que o modelo também gera a capacidade explicativa nesses, com uma performance superior aos modelos *naives* apresentados, e de modelos lineares.

Por fim, a melhor capacidade explicativa dos grandes desvios em relação à média ou mediana é exatamente o que se espera do modelo construído em termos de uso. Com eles, a atividade de controle passa a contar com um instrumento que diferencia grandes desvios explicáveis, daqueles que não o são, gerando um instrumento de identificação rápida e precisa de grandes desvios como anômalos.

3.7 Avaliação no âmbito amostra de teste.

Até agora, todos os resultados foram apresentados com verificações na amostra de validação do *cross validation*. Para a avaliação da performance efetiva, é necessário que os modelos treinados sejam avaliados em amostra segregada, utilizada exclusivamente para fins de testagem. Nesse sentido, anteriormente ao processo de treinamento, os dados foram segregados em 80% para treinamento, e 20% para fim de teste. A partir de agora, os resultados apresentados são os referentes a esta última.

A boa prática de modelagem apontaria que devemos primeiramente decidir qual modelo utilizar com base na performance em *cross validation*, e, posteriormente, reportar os resultados de amostra de teste apenas do modelo escolhido. A intenção desse procedimento é evitar idas e vindas do processo de treinamento, adaptando as escolhas de modelagem para que se adequem a uma amostra de teste específica, o que retiraria a capacidade da validação da amostra de treino indicar uma performance fora dos dados.

Não obstante, para fins de exposição acadêmica dos resultados, entendemos que é pertinente demonstrar as diferenças de resultado fora da amostra do modelo de *Random Forest* treinado e aplicado na integralidade dos dados, do modelo de *Random Forest* treinado e aplicado por quartil de despesa. Da posse dos resultados em *cross validation*, já é possível determinar que a performance geral do modelo por quartil é superior ao geral. Não obstante, entendemos ser de interesse expor as diferenças de comportamento dos modelos na amostra de teste, sem que tais resultados tenham alterado em nada as decisões de modelagem da fase anterior.

3.7.1 Modelo Geral de *Random Forest Regressor* - Amostra de teste.

A avaliação preditiva no âmbito da amostra de teste é imediata, bastando realizar os cálculos das métricas de interesse na comparação entre o resultado predito por aplicação do modelo treinado sobre a amostra de teste e o efetivamente realizado. O único ponto de destaque é que é necessário padronizar os dados da amostra de teste na mesma estaca da de treinamento, utilizando o *StandardScaler* já treinado. Abaixo seguem os resultados.

Tabela 13: *Random Forest Regressor* – Resultados na amostra de teste

Média da amostra de teste	Mediana da amostra de teste	R^2	Erro médio absoluto	Erro mediano absoluto
841.198	19.226	93,33%	303.698	25.792

Fonte: Elaboração própria

Como vemos, os resultados espelham precisamente os encontrados por *cross validation*, indicando ausência de *overfitting* e sendo argumentável pela generalização de sua aplicação no âmbito das despesas dos municípios do Estado de São Paulo (exceto capital).

Verificamos que, na escala absoluta dos dados, o modelo, de fato, apresenta um excelente nível de performance, sendo possível apreender as grandes variações em relação à média. Não obstante, os erros medianos do modelo geral, proporcionalmente às despesas medianas realizadas, são muito significativos.

3.7.2 Modelo *Random Forest Regressor* por quartil (Amostra de teste).

A aplicação preditiva do modelo desenvolvido por quartil, ao contrário do anterior, não é tão imediata. Na medida em que este modelo apenas será útil uma vez determinado em que quartil a despesas a ser predita se encaixa, passa a haver dois caminhos possíveis. De um lado, poderíamos não utilizar os dados das despesas efetivamente realizada, para evitar uma contaminação de informações, o que seria, de uma perspectiva teórica, o mais apropriado.

Nesse sentido, é possível utilizar o modelo de *Random Forest Regressor* geral, treinado na integralidade dos dados e com alto R^2 para prever em que quartil a despesa prevista se encaixaria em função exclusivamente de suas *features*. A partir dessa informação, podemos, então, aplicar o modelo treinado particularmente para aquele quartil, em um refinamento da previsão.

Ocorre que essa estratégia talvez gere erros preditivos desnecessários. Isso porque, no uso do modelo, os auditores já terão acesso às despesas realizadas. O modelo preditivo serve tão somente como elemento de parametrização, referencial. Dessa forma, caso aplicássemos a primeira abordagem, poderíamos carregar os erros preditivos da primeira etapa para a última, eventualmente encaixando a entrada no quartil errado, e aplicando o modelo preditivo não apropriado.

Nesse sentido, projetando a forma de utilização do modelo, optamos por considerar a segunda abordagem. Nesse sentido, pegamos a amostra de teste e segregamos por quartil na escala da amostra de treino. Em seguida, realizamos as predições com o uso de cada modelo apropriado ao quartil, concatenamos os resultados, e avaliamos, por cada uma das métricas abaixo apresentadas.

Tabela 14: *Random Forest Regressor* – Resultados na amostra de teste

Média da amostra de teste	Mediana da amostra de teste	R^2	Erro médio absoluto	Erro mediano absoluto
841.198	19.226	93,56%	257.098	5.292

Fonte: Elaboração própria

Verificamos, pois, uma performance muito significativa. De um lado, há a manutenção de um R^2 acima de 90%, com uma redução muito expressiva do erro mediano absoluto.

3.8 Avaliação final de performance.

Ao fim, verificamos que o mérito do modelo construído é explicar variações muito expressivas das despesas, parametrizando as alterações, de modo que se possa apontar o que são amplas variações esperadas, do que são potenciais anomalias.

As pequenas variações das despesas não são bem capturadas pelo modelo construído, em que pese, ao realizar o treinamento por quartil de despesas, tenhamos diminuído os erros proporcionais nos primeiros quartis. Nesse sentido, entre os modelos desenvolvidos, o de melhor performance foi o *Random Forest Regressor* treinado por quartil de despesas, apresentando um R^2 de 93.56% na amostra de teste.

3.9 Aplicação prática dos modelos desenvolvidos.

É relevante destacar que o modelo desenvolvido já começou a ser utilizado pelo TCE-SP em formato de testes. Em uma avaliação inicial de conceito, os auditores reportaram anomalias identificadas pela utilização do modelo *Random Forest* geral que se mostraram úteis ao processo de auditoria das contas municipais. Nesta seção, apresentaremos alguns desses achados que servem a ilustrar a utilização prática do modelo criado.

A) Erro de Registro

O TCE-SP mostrou que o modelo detectou diversos erros de registro de categorias de despesas por municípios. A título exemplificativo, em determinado município, o valor previsto pelo modelo para gasto pela prefeitura em transporte de servidores e empregados foi de R\$ 39.192,00, ao passo que o valor realizado foi de superior a 4 milhões de reais.

Tal item, portanto, imediatamente foi apontado como uma anomalia que demandou uma análise da auditoria, que verificou que a prefeitura estava erroneamente registrando o

transporte de alunos e escolas nesta categoria. Embora tal anomalia não aparente sustentar um ato lesivo à administração pública, é possível verificar que o modelo preditivo apontou automaticamente um equívoco realizado, útil a ser corrigido pelo processo de auditoria a fim de garantir a higidez dos registros, e com baixíssimo custo operacional. Ademais, alterações de rubricas podem ser veículos para escamotear ilícitos.

B) Anomalias sem explicação aparente.

Para além casos de erros de registros que são imediatamente verificados, o modelo também apontou despesas anômalas que não podem ser facilmente explicadas e, que, em uma avaliação inicial dos auditores, forneciam uma perspectiva relevante à mudança qualitativa do processo de auditoria. Nesse sentido, destacamos aqui dos casos emblemáticos.

No primeiro deles, temos o caso de despesas na categoria sentenças judiciais de pequeno valor. Nesse caso, havia a previsão de gastos de R\$ 575.500,00, ao passo que o valor realizado foi superior a 13 milhões de reais. Diante desse apontamento automático, os auditores abriram as despesas nessa categoria do município e verificaram que, embora nenhuma das despesas individualmente se destacasse, havia uma quantidade de despesas com sentenças judiciais que parecia não se justificar em uma análise inicial, demandando um aprofundamento das causas.

Tal tipo de achado encaixa-se precisamente nos objetivos da construção do modelo preditivo, pois gera uma alteração qualitativa do processo de auditoria, sendo capaz de gerar perspectivas diferente na análise das contas municipais, de tal modo que seja possível não só perceber anomalias a um menor custo operacional, como também verificar anormalidades que possivelmente não seriam verificados no bojo das análises gerais que tradicionalmente realizadas.

Outro caso emblemático foi a de despesa de determinada universidade que realizou a aquisição de um aparelho de processamento de dados por mais de 2 milhões de reais. Tal despesa foi imediatamente apontada pelo modelo como uma despesa anômala, demandando a atenção dos auditores, que efetivamente consideraram a despesa pouco usual.

É claro que tal despesa pode ser uma anomalia positiva, no sentido de que os administradores públicos estão investindo em determinada categoria acima do normalmente praticado pelos demais municípios. Não obstante, é também inegável que esse é um tipo de despesa que requer uma análise mais detida dos auditores, e a utilização de modelo preditivo serviu a um filtro de caso de interesse, diminuindo o custo operacional.

C) Erros preditivos.

Para além dos casos apontados, que constituem exemplos de anomalias que se verificaram claramente úteis ao aumento da performance do processo de auditoria, houve casos de anomalias apontadas que mais indicavam erros preditivos. Para a diminuição desses erros, aprimorando o modelo construído, o TCE-SP passou a disponibilizar em seu sistema de auditoria os valores previstos pelo modelo, ao lado dos efetivamente realizados, com a possibilidade dos auditores filtrarem as despesas pelas discrepâncias encontradas. A partir disso, eles geraram um sistema classificatório, em que os auditores poderão marcar a previsão como adequada ou não. Dessa forma, após esse período de testes, passaremos a contar com uma estrutura classificatória das previsões que poderão servir para um substancial aprimoramento do processo preditivo.

4 Conclusão.

4.1 Contextualização dos resultados.

A presente dissertação tem por objetivo contribuir à pesquisa empírica de métodos de *machine learning* na detecção de corrupção e de atos lesivos à administração de uma forma geral. Conforme demonstrado na revisão bibliográfica, recentes estudos foram desenvolvidos de forma exitosa nos anos recentes nesse sentido.

Não obstante, é relevante marcar que, de uma forma geral, as pesquisas a que tivemos acesso utilizam, majoritariamente, métodos de classificatórios de *machine learning*, em geral tendo por base dados históricos de apontamentos de resultados de auditoria ou de avaliações jurisdicionais passadas. Ocorre que, conforme pontuou Olken (2017), os achados de auditoria demonstram o resultado da incapacidade do agente público de ocultar a corrupção, e não esta per si.

Nesse sentido, os modelos classificatórios de corrupção, tal como vem sendo desenvolvidos, identificam padrões nas características dos achados de corrupção e os generalizam. A alta acurácia das predições se dá *vis a vis* os resultados históricos das atividades de controle. Com estes, é possível estabelecer uma divisão mais apurada dos procedimentos de auditoria, o que tende a aumentar significativamente o resultado global destes, incrementando a probabilidade de que os elementos auditados sejam de atos ilícitos. Seguindo a notação utilizada na seção de introdução, os métodos classificatórios aumentam a p_d , ou seja, a probabilidade de ser submetido a um procedimento de auditoria condicional a ter cometido um ato ilícito.

Não obstante, é preciso marcar que os modelos classificatórios não alteram qualitativamente a análise de auditoria, ou seja, não modificam sensivelmente p_d , a probabilidade de detecção condicional a ter sido submetida a um procedimento de auditoria. Ao contrário, diante da existência de um eventual viés na capacidade de percepção da auditoria, este pode ser acentuado.

Mais concretamente, se existem determinados tipos de atividades que são detectadas recorrentemente com mais dificuldade pelas atividades de auditoria, uma generalização desses resultados também carregará esse viés. Se houver uma redistribuição das atividades de auditoria tomando por base tais resultados, o problema será acentuado, diminuindo ainda mais a capacidade dessas atividades serem descobertas. Tal hipótese pode ser ainda mais preocupante se inserirmos um cenário em que os agentes públicos passam a aprender a existência dessas distorções, e, assim, alterar o comportamento na direção das atividades sob o viés de baixa detecção.

Assim, passa a ser relevante explorar uma abordagem complementar da utilização de modelo de *machine learning* para a detecção de casos de corrupção para que possa haver a alteração qualitativa da atividade de auditoria, de modo a incrementar a probabilidade de detecção condicional ao procedimento de controle. O trabalho empírico que apresentaremos ao longo desta Dissertação foi produzido exatamente neste propósito.

Nesse sentido, ao invés de utilizar as classificações passadas de atividades de controle, é possível pedir que os dados nos contem a sua própria. Implementamos tal estratégia analisando os dados das despesas de 644 municípios do estado de São Paulo entre os anos de 2010 – 2019, utilizando a detalhada base de dados fornecida pela Tribunal de Contas do Estado de São Paulo (TCE-SP).

A detecção de irregularidades de atividades pulverizadas municipais é particularmente desafiadora. Se é verdade que a descentralização administrativa é um instrumento de impacto na ampliação do controle social, que passa a estar mais próxima do centro das decisões relevantes ao interesse local¹⁶, da perspectiva de órgãos de controle, passa a haver uma grande dificuldade de parametrização de um nível de despesa esperado. Nesse cenário, a atividade de controle precisa avaliar atos administrativos realizados por agentes públicos em contextos variados, com distinções demográficas, diferenças de nível e de tipos de despesas e de receitas, e relevantes discrepâncias na capacidade institucional.

¹⁶ (Fisman e Gatti, 1999) argumentam, por análise *cross sectional* de diversos países, que a descentralização diminui o nível de corrupção.

É nesse aspecto que esperamos que o modelo que desenvolvemos possa atuar em complementação com os demais já expostos. O modelo desenvolvido é capaz de capturar as relações não lineares entre características municipais e dos elementos descritivos de despesas. Assim, espera-se que se tenha criado um instrumento de parametrização das despesas municipais, gerando insumos para a ampliação da capacidade de sua análise material.

O modelo preditivo desenvolvido, quando aplicado sobre a integralidade dos dados, gera um R^2 superior a 90% tanto em análise por *cross validation*, quanto na amostra de teste. Tal resultado demonstra que a utilização do modelo é capaz de explicar a quase integralidade da enorme variância de valores de despesas.

No entanto, é preciso expor claramente que, se é verdade que geramos uma capacidade explicativa das variações de grande magnitude em relação à média. As variações menores não são apreendidas pelo mecanismo preditivo. Embora essas últimas sejam proporcionalmente pequenas em relação a uma média largamente enviesada por *outliers*, elas podem ser proporcionalmente muito significativas em relação aos valores realizados, conforme aponta a proporção entre o erro mediano absoluto e a mediana dos gastos.

Para atenuar tais dificuldades, construímos modelos distintos por quartil de despesa. Os resultados dessa abordagem mostram com precisão essa restrição. Como foi apresentado, os modelos preditivos para os três primeiros quartis apresentam R^2 muito inferior ao último.

Diante desses resultados, podemos apontar que o modelo preditivo, de fato, tem uma grande capacidade explicativa para apreender as variações de larga magnitude, e uma baixa capacidade preditiva para as despesas de menor porte. No entanto, ainda que a sua performance seja menor nestes últimos, ainda assim, há uma capacidade explicativa considerável, comportando-se melhor que modelos lineares ou por uma estimação por mediana.

O modelo desenvolvido serve em grande parte ao propósito para o qual ele foi criado, qual seja, servir de parâmetros de comportamentos dos gastos municipais. Na medida em que o modelo é capaz de capturar as grandes variações das despesas municipais, a atividade de controle, passa a contar com um mecanismo automático de diferenciação entre largas

discrepâncias explicáveis (e que, portanto, tendem a não ser anomalias), de grandes variações não previsíveis e que devem suscitar a atenção da auditoria.

Com a adoção do modelo construído, espera-se que haja uma alteração qualitativa do procedimento de controle das contas municipais, permitindo que os auditores tenham em destaque desvios que não seriam evidentes em avaliações individualizadas das contas municipais.

O mecanismo gerará um instrumento de comparação contínua entre os municípios. Assim, não só tende a haver uma diminuição do custo marginal do procedimento de auditoria, que, como vimos na seção de introdução, é um fator determinante do nível ótimo de monitoramento social, e, por consequência, do grau de corrupção, mas também pode a ser possível verificar irregularidades não facilmente detectáveis pelo procedimento tradicional de auditoria.

Nesse sentido, é interessante marcar que tivemos notícia que o modelo desenvolvido começou a ser testado operacionalmente pelas equipes de auditoria do Tribunal de Contas do Estado de São. Em um futuro próximo, poderemos obter retornos quanto ao comportamento do modelo e, certamente, gerar os aprimoramentos relevantes.

4.2 Análise Crítica.

O que todo o histórico de mecanismos desenvolvidos para a detecção de atos de corrupção demonstra é que, não obstante o significativo desenvolvimento de técnicas, esta não é apreensível por um único método, todos apresentando suas limitações.

O modelo apresentado nesta Dissertação não é diferente. Uma primeira crítica evidente é corrupção não é restrita a variações de valor das despesas, podendo derivar de pagamento de propina pela distorção na distribuição dos recursos sem afetar o seu volume. Não obstante este argumento seja verdadeiro, é usualmente argumentado (Mauro, 1996, Olken, 2017) que os agentes públicos corruptos geram distorções na alocação das categorias de recursos em favor daquelas que oferecem maiores possibilidades de desvios, o que, a depender da intensidade da distorção, poderia ser marcado em nosso modelo.

Outro ponto que deve ser arguido é que a corrupção não demanda qualquer relação com as despesas públicas, podendo, por exemplo, decorrer das atividades regulatórias, das atividades de monitoramento do Estado. Quanto a esta crítica, não há contra-argumentos. De fato, nosso modelo não é capaz de detectar atividades de corrupção não relacionadas a despesas públicas. Vale ressaltar, contudo, que o volume de corrupção no âmbito destas é enorme, conforme os recentes eventos na própria história brasileira demonstram, de modo que o desenvolvimento de instrumentos que sejam capazes de atacar esse aspecto do fenômeno da corrupção já seria de grande valia.

De outra parte, uma crítica muito mais sensível ao modelo proposto seria a de que ele não captura atos de corrupção no âmbito das despesas públicas, mas sim distorções a padrões desenvolvidos pelos próprios municípios.

De fato, o modelo identifica distorções, sejam positivas, sejam negativas. Apenas para fins exemplificativos, se os municípios fossem corruptos de forma uniforme, com a mesma intensidade e no âmbito das mesmas categorias de despesas, tais atos de corrupção não seriam percebidos. Se houvesse um choque de gestão em determinado município, com uma redistribuição das despesas, por exemplo, reduzindo significativamente os gastos com combustíveis e redistribuindo para tecnologia de educação infantil, tal aspecto será considerado uma distorção tal como se o inverso houvesse acontecido.

Por fim, é preciso frisar que, em que pese a motivação para construção do modelo preditivo tenha sido contribuir com a recente produção empírica de mecanismos de detecção de corrupção, os atos marcados como anômalos não necessariamente derivam desse tipo de ato, mesmo em sua ampla definição como: *“a violação de uma regra por um agente público para auferir benefícios para si ou para outrem”*.

Os desvios relevantes encontrados pela utilização do modelo preditivo construído, para além de erros do modelo, podem simplesmente derivar de fatores idiossincráticos dos municípios, do mercado de bens e serviços locais, da restrição técnica da burocracia local, ou de inúmeros outros fatores. Não obstante tais desvios ainda não configurem atos lesivos

dolosos contra a administração pública, muitas das vezes, devem ser levados à atenção da auditoria, para que, eventualmente, possam ser induzidos aprimoramentos institucionais.

Não há dúvidas que o modelo proposto apresenta limitações, como era de se esperar. Não se trata de uma solução para substituir a atividade de auditoria, mas sim para criar um instrumento que gere parâmetros materiais de análise de anomalias de despesas com o intuito de ampliar a capacidade de monitoramento dos procedimentos de controle e, dessa forma, aumentar a probabilidade de detecção de atos lesivos à administração pública.

Não obstante as limitações apresentadas, dentro desse escopo pretendido, acreditamos que esta Dissertação gerou uma contribuição relevante.

4.3 Pesquisa Futura e Aprimoramentos.

É possível apontar caminhos para o aprimoramento do modelo apresentado. A primeira abordagem para o incremento de performance seria a melhoria dos dados para o treinamento do modelo preditivo. Nesse sentido, já destacamos que não obtivemos variações anuais de diversas características municipais, tais como PIB, população, número de servidores e número de alunos. Ademais, poderiam ser testadas outras características que podem ser determinantes em aumentar a capacidade preditiva, tais como IDH, GINI, distribuição etária, dados acerca do sistema de saúde local, preços e salários locais, dentre outros.

De outra parte, poderiam ser exploradas outras estratégias de agrupamento dos dados de despesas a partir da base da AUDESP. Nesse sentido, poderia ser avaliado como o uso da classificação das despesas pelos atributos *ds_funcao_governo* ou *ds_subfuncao_governo* impactariam não só a performance preditiva, mas a qualidade da informação provida ao processo de auditoria. Ademais, diante de maior capacidade computacional, poderia ser explorado o uso de informações sobre o tipo de procedimento de licitação utilizado, e de fonte dos recursos como fatores preditivos. Seria também de interesse aumentar o grau de granularidade dos dados, ao menos ao nível mensal, com o objetivo de aferir determinações de sazonalidade no comportamento dos gastos.

Um outro caminho relevante de aprimoramento é a expansão regional do modelo desenvolvido, de modo a tentar avaliar dados de outros estados brasileiros. Acreditamos que há um relativo grau de padronização dos sistemas de controle das contas municipais, de modo que, a partir de necessárias interações institucionais, poderia ser criado um sistema nacional que buscasse aumentar a comparabilidade dos gastos municipais.

Por fim, podem ser avaliados outros algoritmos de *machine learning*, incluindo aqui um teste mais amplo de hiperparâmetros dos modelos. Em particular, os modelos de redes neurais foram testados apenas em relativa superficialidade, restrita a redes sequenciais densas, diante de uma menor familiaridade com o campo de *deep learning*. Acreditamos, portanto, que este pode ser um relevante campo de aprofundamento da pesquisa desenvolvida.

REFERÊNCIAS

ASH, Elliott; GALLETTA, Sergio; e GIOMMONI, Tommaso. **A Machine Learning Approach to Analyze and Support Anti-Corruption Policy**. Abril de 2020, 34p.

AVIS, Eric; FERRAZ, Claudio Ferraz; e FINAN, Frederico. **Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians**. Sem data. 50p.

BANERJEE, Abhijit; MULLAINATHAN Sendhil; e HANNA Rema. **Corruption**. Nber Working Paper Series, Abril de 2012.

BECKER, Gary S.; e STIGLER, George J. **Law Enforcement, Malfeasance, and Compensation of Enforcers**. *The Journal of Legal Studies* 3, 1–18p. <https://doi.org/10.1086/467507>

DEININGER, Klaus; e Mpuga, Paul. **Does Greater Accountability Improve the Quality of Public Service Delivery? Evidence from Uganda**. *World Development* 33, 171–91p. <https://doi.org/10.1016/j.worlddev.2004.09.002>.

DI TELLA, Rafael; e Schargrodsky, Ernesto. **The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires**. *The Journal of Law and Economics* 46, 269–92p. <https://doi.org/10.1086/345578>.

FISMAN, Raymond; e GATTI, Roberta. **Decentralization and Corruption: Evidence Across Countries?** Sem data, 19p.

GALLEGO, Jorge; RIVERO Gonzalo; e MARTÍNEZ, Juan. **Preventing Rather than Punishing: An Early Warning Model of Malfeasance in Public Procurement**. *International Journal of Forecasting* 37, 360–77p. <https://doi.org/10.1016/j.ijforecast.2020.06.006>.

GRACE, Emily; RAI, Ankit; REDMILES, Elissa; e GHANI, Rayid. **Detecting Fraud, Corruption, and Collusion in International Development Contracts: The Design of a Proof-of-Concept Automated System**. *IEEE International Conference on Big Data (Big*

Data), Washington DC, USA: IEEE, 2016, 1444–1553p.
<https://doi.org/10.1109/BigData.2016.7840752>.

HUNT, Jennifer. **How corruption hits people when they are down**. *Nber Working Paper No. 12490*, Agosto de 2006. <http://www.nber.org/papers/w12490>.

MAURO, Paolo. **The Effects of Corruption on Growth, Investment, and Public Expenditure: A Cross Country Analysis**. *IMF Working Paper No. 96/98*, 1996, 25p.

NIEHAUS, Paul; e SUKHTANKAR, Sandip. **The Marginal Rate of Corruption in Public Programs: Evidence from India**. *Journal of Public Economics* 104, 52–64p.
<https://doi.org/10.1016/j.jpubeco.2013.05.001>.

OLKEN, Benjamin A. **Monitoring Corruption: Evidence from a Field Experiment in Indonesia**. *Journal of Political Economy*, 2007, 50p.

OLKEN, Benjamin A.; e PANDE, Rohini. **Corruption in Developing Countries». Annual Review of Economics** 4, n.º. 1. Setembro de 2012, : 479–509p. <https://doi.org/10.1146/annurev-economics-080511-110917>.

PREM, Mounu. **What Predicts Corruption?** 17 de Abril de 2020, 34p.

RIJCKEGHEM, Caroline; e WEDER, Beatrice. **Corruption and the rate of temptation: Do low wages in the civil service cause corruption?** *IMF Working Paper*, Junho de 1997.

SHLEIFER, Andrei; e VISHNY, Robert W. **Corruption**. *The Quarterly Journal of Economics* 108, n. 3, 599–617p.

VARGAS, Juan F. **Corruption in the Times of Pandemia**. Sem data, 47p.

ZAMBONI, Yves; e LITSCHIG, Stephan. **Audit Risk and Rent Extraction: Evidence from a Randomized Evaluation in Brazil**. *Journal of Development Economics* 134, 133–149p.
<https://doi.org/10.1016/j.jdeveco.2018.03.008>